

OM of: US-09-030-606-174 to: PIR_60:* out_format : pfs
Date: Sep 25, 1999 6:28 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171617_29825/app-query.fasta.1
-DB=PIR_60 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-030-606-174
Query length: 1459
Database: PIR_60:*
Database sequences: 122810
Database length: 40065486
Search time (sec): 182.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|-------------|------|--------|--------|---------|-----|---------------------------------|
| pir4:C40201 | - | 315.00 | 466.44 | 4.7e-19 | 613 | artifact-warning sequence (tran |
| pir4:F40201 | - | 289.50 | 426.55 | 7.1e-17 | 673 | artifact-warning sequence (tran |
| pir4:A40201 | - | 285.00 | 420.25 | 1.7e-16 | 627 | artifact-warning sequence (tran |
| pir4:F40201 | + | 283.00 | 416.59 | 2.5e-16 | 673 | artifact-warning sequence (tran |
| pir4:C40201 | + | 271.00 | 398.98 | 2.7e-15 | 613 | artifact-warning sequence (tran |
| pir4:A40201 | + | 247.50 | 362.76 | 2.7e-13 | 627 | artifact-warning sequence (tran |
| pir4:A46010 | - | 246.50 | 376.76 | 2.8e-13 | 100 | X-linked retinopathy protein (C |
| pir2:I38022 | + | 242.00 | 364.17 | 7.3e-13 | 196 | hypothetical protein - human |
| pir2:UQ1472 | + | 237.50 | 355.35 | 1.8e-12 | 246 | trypsin (EC 3.4.21.4) V precurs |
| pir2:UQ1471 | + | 237.50 | 355.35 | 1.8e-12 | 246 | trypsin (EC 3.4.21.4) V precurs |
| pir1:TRDGC | + | 236.00 | 353.05 | 2.4e-12 | 246 | trypsin (EC 3.4.21.4) precurs |
| pir4:E40201 | - | 233.50 | 341.71 | 4.2e-12 | 597 | artifact-warning sequence (tran |
| pir2:S05494 | + | 233.00 | 348.41 | 4.3e-12 | 247 | trypsin (EC 3.4.21.4) IV precu |
| pir1:TRBOTR | + | 226.50 | 339.09 | 1.5e-11 | 229 | trypsin (EC 3.4.21.4) precursor |
| pir2:A27547 | + | 224.50 | 335.38 | 2.3e-11 | 247 | trypsin (EC 3.4.21.4) precursor |
| pir2:A53968 | + | 222.00 | 331.34 | 3.8e-11 | 253 | serine proteinase SCCE precurs |
| pir1:A25852 | + | 221.50 | 330.78 | 4.2e-11 | 247 | trypsin (EC 3.4.21.4) I precurs |
| pir2:S49489 | + | 221.00 | 330.19 | 4.6e-11 | 242 | trypsin (EC 3.4.21.4) precursor |
| pir2:S31779 | + | 220.50 | 329.56 | 5.1e-11 | 238 | trypsin (EC 3.4.21.4) III precu |
| pir2:S55065 | + | 220.50 | 329.21 | 5.1e-11 | 248 | trypsin (EC 3.4.21.4) I precurs |
| pir2:S55067 | + | 219.00 | 326.91 | 6.8e-11 | 248 | trypsin (EC 3.4.21.4) I precurs |
| pir2:S55066 | + | 218.50 | 326.75 | 7.5e-11 | 231 | trypsin (EC 3.4.21.4) II precu |
| pir2:S31778 | + | 213.50 | 319.69 | 2.0e-10 | 215 | trypsin (EC 3.4.21.4) IB - Ala |
| pir2:S66661 | + | 213.50 | 318.69 | 2.0e-10 | 242 | trypsin (EC 3.4.21.4) I precurs |
| pir2:S31775 | + | 213.50 | 318.69 | 2.0e-10 | 242 | trypsin (EC 3.4.21.4) IA precu |
| pir2:S31776 | + | 211.50 | 315.59 | 3.0e-10 | 243 | trypsin (EC 3.4.21.4) precursor |
| pir2:A35871 | + | 211.50 | 315.05 | 3.0e-10 | 259 | trypsin (EC 3.4.21.4) IV form b |
| pir2:I38363 | + | 211.50 | 313.69 | 3.0e-10 | 304 | trypsin (EC 3.4.21.4) IV form a |
| pir2:S33496 | + | 209.00 | 311.51 | 4.9e-10 | 250 | trypsin (EC 3.4.21.4) - plate |
| pir2:T01779 | + | 208.50 | 310.88 | 5.4e-10 | 246 | trypsin (EC 3.4.21.4) I precurs |
| pir1:TRRT1 | + | 208.00 | 310.12 | 5.9e-10 | 246 | trypsin (EC 3.4.21.4) precursor |
| pir2:B25528 | + | 207.50 | 309.88 | 6.5e-10 | 231 | trypsin (EC 3.4.21.4) precursor |
| pir1:TRPGTR | + | 207.00 | 309.19 | 7.2e-10 | 229 | trypsin (EC 3.4.21.4) precursor |
| pir1:B25852 | + | 206.50 | 307.78 | 8.0e-10 | 247 | trypsin (EC 3.4.21.4) II precu |
| pir2:B31136 | + | 206.00 | 306.61 | 8.8e-10 | 259 | tissue kallikrein (EC 3.4.21.35 |
| pir2:I56559 | + | 205.50 | 305.82 | 9.7e-10 | 260 | neuropsin - mouse |
| pir4:E40201 | + | 202.50 | 294.18 | 1.9e-09 | 597 | artifact-warning sequence (tran |
| pir1:TRRT2 | + | 202.00 | 300.92 | 1.9e-09 | 246 | trypsin (EC 3.4.21.4) II precu |
| pir2:S01971 | + | 202.00 | 300.42 | 1.9e-09 | 261 | tissue kallikrein (EC 3.4.21.35 |
| pir2:S15685 | + | 199.50 | 296.95 | 3.2e-09 | 250 | kallikrein, glandular - multima |
| pir2:S13813 | + | 197.00 | 293.22 | 5.2e-09 | 247 | trypsin (EC 3.4.21.4) - bovine |

pir2:S39048 + 196.50 292.66 5.7e-09 241 | trypsin (EC 3.4.21.4) X - At
pir2:S39047 + 195.50 291.16 6.9e-09 240 | trypsin (EC 3.4.21.4) I - At
pir1:TRDG + 195.00 290.15 7.6e-09 247 | trypsin (EC 3.4.21.4) precu

seq_name: pir4:C40201

seq_documentation_block:
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Contents: conceptual translations of ALU-repeat sequences
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLAS>
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potenti
A;Reference number: A40200; MUID:92241891
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of

alignment_scores:
Quality: 315.00 Length: 126
Ratio: 3.706 Gaps: 4
Percent Similarity: 67.460 Percent Identity: 57.143

alignment_block:
US-09-030-606-174/rev x C40201 ..

Align seg 1/1 to: C40201 from: 1 to: 613

1459 TTTT TTTT TTTT TTTT TTTGAGACAGAGTCTTACTGTGTGCCCCAGCTGG 1410
|||||
310 PhePhePhePhePhePheGluThrGluSerHisSerValThrGlnAlaG1 326
1409 AGTATAGTGTGTGATCTCAACTCAACTCAACTCTGCTGCCCATATTTCA 1360
|||||
326 YValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheM 343
1359 AGCAATTCCTGCTCCAGCTCCCAAGTAGCTGGATTACAGCGCCTG 1310
|||||
343 etProPheSerCysLeuSerLeuLeuArgThrTrpAspTyrArgArg... 358
1309 CCACCATATCCAGCTAAC.....TTTGTATTTTGTAGTACAGACAGGAT 1266
|||||
359 ..ProHisHisAlaGlnLeuIlePheCysIlePheSerArgAsnGlyVa 374
1265 TTCACCATTTTGGCCAGGCTGCTTGAACCTCTTACCTCAAGTGATCTG 1216
: ||| |||||
374 IleuProCysTrpProGlyTyrPserArgThrProGlyLeuMetIleHis. 390
1215 CCTGCCTCGCCCTCCCAAGTGTGGATTACAGGATGAGCCACCTGCC 1166
||| |||||
391 ProProLeuProProLysValLeuGlyLeuGlnAla**AlaThrAlaPr 407
1165 CAGCCTCTTGTAGAGTCTTGATTTTCTGTCTCTCTGTTTTCACC 1116
| |||||
407 cArg.....PhePhePhePhePhe*****L 418
1115 TCTATGATCTGTGCTCACTGTTCCC 1090
|| :|||
418 euArgGlnSerLeuThrLeuSerPro 426
seq_name: pir4:F40201

seq_documentation_block:
artifact-warning sequence (translated Alu class F) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: F40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Contents: conceptual translations of Alu-repeat sequences
A:Accession: F40201
A:Molecule type: DNA
A:Residues: 1-673 <CLAF>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

alignment_scores:
Quality: 289.50 Length: 347
Ratio: 1.713 Gaps: 17
Percent Similarity: 48.703 Percent Identity: 32.853

alignment_block:

US-09-030-606-174/rev x F40201 ..

Align seg 1/1 to: F40201 from: 1 to: 673

1459 TTTT TTTT TTTT TTTT TTGAGACAGAGTCTTACTC 1425
||||| ||||||| ||||||| |||||||
345 PhePhePhePhePhePheCysPheValPheValPheGluValGlyProCysSe 361
1424 TGTGCCCCAGCTGGAGTATAGTGGTGTGATCTCAACTCACTCACTC 1375
| : : : ||||| ||||| ||||| : : : ||||| : : : |||||
361 rAspThrProAlaGlyValGlnTrp.GlnValLeuAlaHisCysSerLeu 377
1374 TGCCTCCCATATTCAGCAATTCCTGCTGCTCAGCCTCCCAAGTAGCTGG 1325
||| ||||| : : : ||||| : : : ||||| : : : |||||
378 AsnLeuLeuGlySerSerAspSerProAlaSerValSerArgValAlaG1 394
1324 GATTACAGGCGCTGCCACCATATCCAGCTAACTTTTGTATTTTAGTAC 1275
||||| ||||| : : : ||||| ||||| : : : ||||| : : : |||||
394 YIleThrGlyMetClyArgHisSerTrpLeuIleTyrValPheLeuIleG 411
1274 AGACAGATTTGACCATTTTGGCCAGGCTGGTCTTGAACCTCCTTACCTCA 1225
: : ||||| ||||| ||||| ||||| : : ||||| : : |||||
411 lUThrGlnPheHisValAspGlnAlaGlyLeuLysLeuLeuThrSer 427
1224 AGTATCTGCTGCTGCTGCTGCCCAAGTCTGGGATTACAGCATGAGC 1175
||||| ||||| : : ||||| ||||| ||||| |||||
428 SerAspLeuProSerTrpSerProLysValLeuGly**GlnAla**Al 444
1174 CACCCTGCCACG.....CTCTTTGTAGAGTCTTGATTTTCTGTCTTC 1131
||||| ||||| ||||| : : ||||| : : ||||| : : |||||
444 aThrThrProSer*****PheValPheGlyGlyPhePhePhePheP 461
1130 TCTTGTGTTTCACTCTAT.....GAATCTGTGTCACTGTTTCCCTCT 1087
: : ||||| : : ||||| : : ||||| : : |||||
461 hePheAlaLeuPheLeuPheLeuArg**AlaLeuAlaLeuThrProArg 477
1086 GGGTACACAGTTGACCTTGACACACCTGTGTTGATGCACAGTCCAC 1037
478LeuGluCysSerG1 482
1036 TTATCTGTGATTTTTCATATAACACATCAGAAAAATTTAGAGAGATT 987
||||| : : |||||
482 yLysPheTrpLeuThrAlaAlaSerThrSerTrp..... 493

986 GCAACAATTAAAAAATTCACAGACGAACTGTGTAGCTTAGAATATATCA 937
: : : : ||||| : : : :
494ValGlnAlaIleLeuLeuProLeuSerProVal**LeuGly 507
936 AAAAATTCACAAGAGTATATCATGATGCATAAACGTATGCATTAAT 887
507 507
886 CGACTATTATGTTATGTTAGAGCTCCCGCTCAACAGTAGGCTATTCT 837
||| ||| : : ||||| : : |||||
508LeuGlnAlaTrpAlaAlaIleProGly**PheMetTyrPhe. 521
836 AGTCAGGTTTGGGAGTCAAAAGTTATAAGA..... 804
: : : : : : : : : :
522 ****ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer** 538
803GGATTTTCTATTTGGGTGG.....A 785
538 *ProGlnValIleCysProProGlyLeuProLysCysTrpAspArgA 555
784 GGTACCG.....CCCTCAGGCCCATGCTGTGGAAGTCAACTGCAT 741
||| : : ||||| : : ||||| : : |||||
555 rgluProProHisProAla*****LeuPheLeuGlyValPhePhe 571
740 TTTATGTGTGTGTGTTTCTGCTCCCATGCTGTGTTCTGTCTCTCT 691
||| : : ||||| : : ||||| : : |||||
572 PhePhe.....PheLeu.....LeuCysPheCysPhe**As 582
690 TCCTTTGCTTCTCTTATTCCTGTGTTTATTTCTGTGTTCTCTCAGTTT 641
||||| ||||| : : : : : : : :
582 PargProLeuLeu**HisProGlyTrpSerAlaValAlaSerPheGlys 599
640 CTCTCTCT.....AGTTGTCACCTGTCTC 618
||||| : : ||||| ||||| |||||
599 erLeuGlnProGlnProProGlyPheLysArgPheSerCysLeuCysLeu 615
617 CCTGTGTCTCTCCATCTCTGCATCTCGCCAT.....GC 583
||| : : ||||| ||||| |||||
616 ProCysSerTrpAspTyr.....ArgHisGlyProProPheLeuAl 629
582 GGCCCTGTGTGTC.....TCTGTCTCCCC 558
| ||||| : : ||||| : : |||||
629 aAsnLeuCysIlePheAsnArgAspThrValSerPro 641

seq_name: pir4:A40201

seq_documentation_block:
artifact-warning sequence (translated Alu class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Contents: conceptual translations of Alu-repeat sequences
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLAF>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti
A:Reference number: A40200; MUID:92241891
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

alignment_scores:
Quality: 285.00 Length: 125
Ratio: 3.476 Gaps: 4
Percent Similarity: 65.600 Percent Identity: 56.000

alignment_block:

US-09-030-606-174/rev x A40201 ..

Align seg 1/1 to: A40201 from: 1 to: 627

```
1456 TTTT TTTT TTTT TTTGAGACAGAGTCTACTCTGTTGCCACAGCTGGAGT 1407
      |||::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
321 PheLeuPhePheSerGluThrGluSerArgSerValAlaLysAlaGlyVa 337
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1406 ATAGTGTGATCTCAACTCACTCAACCTCTGCTCCATATTCAAGC 1357
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
337 IGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheThrP 354
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1356 AATTCCTCGCTCAGCTCCCAAGTAGCTGGGATTACAGCGCTGCCA 1307
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
354 rPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArg..... 368
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1306 CCATATCCAGCTAAC.....TTTGTATTTTGTAGTACAGACAGATTC 1263
      |||::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
369 ProHisTyrAlaArgLeuIlePheCysIlePheSerArgAspGlyValSe 385
      |||::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1262 ACCATTTTGGCCAGGCTGGTCTTGAACCTCTTACCTCAAGTATGCTG 1213
      |||::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
385 rProTrp**ProGlyTrpSerArgSerProAspLeu.ValIleArgPro 401
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1212 GCCTCGCCTCCCAAGTCTGGGATTACAGCGCATGACCCCTGCCAG 1163
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
402 ProTrpProProLysValLeuGlyLeuGlnAla**AspArgAlaArg** 418
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1162 CCTCTTTGTAGAGTCTGATTTTTCCTGTTCTCTCTGTTTACCTCT 1113
      :  |||  |||  |||  |||  |||  |||  |||  |||  |||
418 *****.....PhePheIlePheCysPheTyrPhePheProA 431
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1112 ATGAATCTGTGTACTGTTTCCC 1090
      ::|||::|||  |||  |||
431 rArgSerLeuAlaLeuSerPro 438
```

seq_name: plr4:F40201

seq_documentation_block:

artifact-warning sequence (translated ALU class F) - human
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C/Accession: F40201
R/Claverie, J.M.
personal communication, 1992
A/Reference number: A40201
A/Contents: conceptual translations of ALU-repeat sequences
A/Accession: F40201
A/Molecule type: DNA
A/Residues: 1-673 <CLA>
R/Claverie, J.M.
Genomics 12, 838-841, 1992
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential
A/Reference number: A40200; MUID:92241891
A/Contents: annotation
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C/Comment: Any significant similarity of a predicted protein sequence to a portion of th

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 283.00 | Length: | 104 |
| Ratio: | 3.724 | Gaps: | 2 |
| Percent Similarity: | 73.077 | Percent Identity: | 64.423 |

alignment_block:

US-09-030-606-174 x F40201 ..

Align seg 1/1 to: F40201 from: 1 to: 673

1161 GCCTGGCAGGCTGCTCATGCTGTAAATCCAGACACTTGGGAGCGAG 1210

||||::| ||||| ||| ||| ||| ||| ||| ||| ||| |||

1 GlyTrpValTrpTrpLeuThrProValIleProAlaLeuTrpGluThrAr 17

1211 GCAGCAGATCACTTGAGTAAAGAGTTCAGACACAGCCTGGCCAAATG 1260

:::||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

17 rArgAlaAspHisLeuArgSerGlyVal**AspGlnProAspGlnHisG 34

1261 GTCAATCTCTGTCTACTAAATAATACAAAAGTTAGCTGGATATGTTGCC 1310

|||| ::|||::| ||| ||| ||| ||| ||| ||| ||| |||

34 IyGluThrValSerLeuLeuLysIleHisLysLeuAlaArgAsnGlyGly 50

1311 AGCGCCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGAGCAATTGCTT 1360

||| ||| ||| ||| ||| ||| ||| ||| ||| |||

51 ProCysLeu**SerGlnLeuHisGlyArgGlnArgGlnGlnAsnArgLe 67

1361 GAATATGGAGCAGAGGTTGAAGTGAAGTGAATCAGACCACTATACTC 1410

|||| ||| ||| ||| ||| ||| ||| ||| ||| |||

67 uAsnProGlyGly**GlyCysSerGluProLysLeu.AlaThrAlaLeu 83

1411 CAGCTGGGCGCAACAGAGTAAGACTCTGTCTCAAAA.....AA 1448

||| ||| ||| ||| ||| ||| ||| ||| ||| |||

84 GlnProGlyCysGlnSerLysGlyLeuSerGlnLysGlnLysGlnSerLy 100

1449 AAAAAAAAAA 1458

||||| |||

100 sLysLysLys 103

seq_name: plr4:C40201

seq_documentation_block:

artifact-warning sequence (translated ALU class C) - human
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
R/Claverie, J.M.
personal communication, 1992
A/Reference number: A40201
A/Contents: conceptual translations of ALU-repeat sequences
A/Accession: C40201
A/Molecule type: DNA
A/Residues: 1-613 <CLA>
R/Claverie, J.M.
Genomics 12, 838-841, 1992
A/Title: Identifying coding exons by similarity search: Alu-derived and other potenti
A/Reference number: A40200; MUID:92241891
A/Contents: annotation
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'X'.
C/Comment: Any significant similarity of a predicted protein sequence to a portion of

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 271.00 | Length: | 110 |
| Ratio: | 3.346 | Gaps: | 1 |
| Percent Similarity: | 73.636 | Percent Identity: | 55.455 |

alignment_block:

US-09-030-606-174 x C40201 ..

Align seg 1/1 to: C40201 from: 1 to: 613

1132 AAACAGGAAAAATCAAGACTCTACAAAGAGGCTGGGCGAGGCTCATG 1181

:::||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

94 GlnLysLysLysLysLysLys*****AlaGlyCysSerGlySerAr 110

1182 CCTGAATCCAGCACTTTGGGAGGCGAGGCGAGGAGATCACTTGAGGTA 1231

||| ||| ||| ||| ||| ||| ||| ||| ||| |||

110 gLeu**SerGlnHisPheGlyArg.GlnArgTrpValAspHisGluAla 126

1232 AGGAGTTCAAGACCAAGCTGGCCAAAATGTTGAATCTGTCTGTACTAA 1281

||||| ||| ||| ||| ||| ||| ||| ||| ||| |||


```
127 ArgSerSerArgProAlaTrpProThrTrpGlnAsnProIleSerThrLy 143
1282 AAATACAAAGTTAGCTGATATGCTGGCAGCGCTGTAATCCAGCTA 1331
      |||||
      |||||
143 sAsnThrLysAsn**LeuGlyMetValAlaGAlaProValProArgT 160
      |||||
1332 CTGGAGGCTGAGCAGAGAGAAATTGCTGATATGGAGCAGAGTGTG 1381
      |||||
      |||||
160 hrGlnLysAlaGlnAlaGlnGlnTrpHisLysProGlnArgTrpSerLeu 176
      |||||
1382 AAGTAGTTGAGATCACACCCTATCTCCAGCTGGGGCACAAGAGTAAG 1431
      |||||
      |||||
177 Gln**AlaGlnLysSerProLeuHisSerSerLeuGlyAspArgValAr 193
      |||||
1432 ACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
      |||||
193 gLeuCysLeuLysLysLysLysLysLys 202
seq_name: p1r4:A40201
```

```
seq_documentation_block:
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Contents: conceptual translations of ALU-repeat sequences
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th
```

```
alignment_scores:
Quality: 247.50      Length: 98
Ratio: 3.300        Gaps: 2
Percent Similarity: 76.531      Percent Identity: 61.224
```

alignment_block:

US-09-030-606-174 x A40201 ..

Align seg 1/1 to: A40201 from: 1 to: 627

```
1166 GGCAGGGTG...GCTCATGCCCTGTAATCCAGCAGCACTTTGGAGCGGAGGC 1212
      |||||
      |||||
1      GLyArgAlaArgSerHisAlaCysAsnProSerThrLeuGlyGlnGln 17
1213 AGGCAGATCACTTGAGGTAAGGAGTTCAGACCAAGCCCTGGCCAAATGGT 1262
      |||||
      |||||
17      yGlyArgGlnLeuThrArgSerGlyAspArgAspHisProGlyTyrHisGly 33
1263 GAAATCCTGTCTGTAATAAAATACAAAGTTAGCTGATATGGTGGCAG 1312
      |||||
      |||||
34      GluThrProSerLeuLeuLysLysLysLysLysSerArgAla**CysGln 50
1313 GCGCCTGTAATCCAGCTACTTGGAGGCTGAGCAGAGAGATTCCTTGA 1362
      |||||
      |||||
50      yArgLeu**SerGlnLeuLeuGlyArgLeuArgGlnGlnGlnGlnGlyVala 67
1363 ATATGGAGGAGAGAGGTGAGAGTGAAGTTGAGATCACACCCTATCTCCA 1412
      |||||
      |||||
67      snProGlyGlyGlyAlaCysSerGlnProArgSerArgHisCysThrPro 83
```

```
1413 GCTGGGCAACAGAGTAGACTCTGTCTCAAAAAAAAAAAAA 1454
      |||||
      |||||
84      AlaLeuAlaThrGlnArgAspSerValSerGlnLysAsnLys 97
seq_name: p1r2:A46010
```

```
seq_documentation_block:
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: A46010
R:Wong, P.; Macdonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A:Title: Identification and partial characterization of a candidate gene for X-linked
A:Reference number: A46010; MUID:93224131
A:Accession: A46010
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-100 <WON>
A:Cross-references: GB:S58722; NID:g299470; PID:g299471
A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIPI:129340)
```

```
alignment_scores:
Quality: 246.50      Length: 117
Ratio: 3.160        Gaps: 4
Percent Similarity: 66.667      Percent Identity: 51.282
```

alignment_block:

US-09-030-606-174/rev x A46010 ..

Align seg 1/1 to: A46010 from: 1 to: 100

```
1459 TTTTNTTTTNTTTTNTTTTGGACAGAGAGTCTTACTCTGTGCCCCAGCTGG 1410
      |||||
      |||||
1      PhePhePhePhePhePheGlnThrGlnSerCysSerValAlaGlnAlaGln 17
1409 AGTATAGTGGTGTGATCTCAACTCACTTCAACCTCTGCCCTCCATATTC 1360
      |||||
      |||||
17      yValGlnTrpCysAspLeuGlySerLeuLysSerProPro.ProGlySer 33
1359 AGCAATTCCTCGCTCAGCCTCCCAAGTAGCTGGGATTACAGCGCGCTG 1310
      |||||
      |||||
34      SerAspSerProAlaSerAlaSerArgValAlaGlyLeuThrGlyMetHis 50
1309 CCACCATATCCAGCTAAGTTTGTATTTTTAGTACAGACAGGATTTTCAC 1261
      |||||
      |||||
50      sHisHisThrGlnLeuLeuPheValPheLeuValGlnThrGlySerHisM 67
1260 ..CATTTGGCCAGGCGCTGTGTAAGTCTTGAAGTCTTCACTCTCAAGTGAT..... 1219
      |||||
      |||||
67      etGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAsnAlaLys 83
1218 CTGCCCTGCGCTCGCCTCCCAAGTGCTGGGATTACAGGATGAGCCACCTT 1169
      |||||
      |||||
84      IleThrAlaArgAlaProArgAspLeu..... 92
1168 GCCCAGGCTCTTTGTAGAGTCTGATTTTTCCTGTTCTCTGCTGTTT 1120
      |||||
      |||||
93      .....PhePhePhePhePhePhePhePhePhe 100
seq_name: p1r2:I38022
seq_documentation_block:
hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 01-Nov-1996
C:Accession: I38022
R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zelisstra-Ryalls, J.H.; Somerville, R
Cancer Res. 50, 5658-5667, 1990
A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both r
A:Reference number: I38021
A:Accession: I38022
```


A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: EMBL:X55777; NID:g288143; PID:g288145

alignment_scores:
Quality: 242.00 Length: 78
Ratio: 3.841 Gaps: 0
Percent Similarity: 80.769 Percent Identity: 69.231

alignment_block:
US-09-030-606-174 x I38022 ..

Align seg 1/1 to: I38022 from: 1 to: 196

```
1221 CACTTGAGTGAAGAGTTCAAGACACCGCTGGCCAAATGTTGAATCCT 1270
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 HisLeuArgSerGlyValGlnAspTyrProGlnHisGlyLysIlePr 19
1271 GTCTGTACTAAAAATACAAAGTTAGCTGATATGCTGGCAGCGCCTGT 1320
|||:::||||| |||||:::||||| ||||| ||||| ||||| |||||
19 oSerLeuLeuLysIleGlnGlnLeuAlaGlyHisGlyLysIlePr 36
1321 AATCCAGCTACTTGGGAGGCTGAGGCAGACAATGCTTGAATATGGGA 1370
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
36 InSerGlnLeuLeuArgArgLeuArgGlnGlnAsnHisLeuAsnSerGly 52
1371 GGCAGAGGTTGAAGTGAAGTTGAGATCACACCACTATCTCCAGCTGGGGC 1420
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 GlyArgGlyCysSerGlnProLysSerHisLeuLysIleProAlaTrpVa 69
1421 AACAGAGTAAGACTCTGTCTCAAAAAAATAAAAAA 1454
:||||| :||||| ||||| ||||| ||||| ||||| ||||| |||||
69 lThrGlnGlyAspSerValSerLysGlnAsnLys 80
```

seq_name: pir2:JQ1472

seq_documentation_block:

trypsin (EC 3.4.21.4) V precursor, b-form - rat
C:Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
C;Accession: JQ1472
R;Kang, J.; Wiegand, U.; Mueller-Hill, B.
Gene 110, 181-187, 1992
A;Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
A;Reference number: JQ1471; MUID:92165057
A;Accession: JQ1472
A:Molecule type: mRNA
A;Residues: 1-246 <KAN>
A;Cross-references: EMBL:X59013; NID:g57414; PID:g57415
A;Experimental source: pancreas
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; protein digestion; serine proteinase; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-24/Domain: activation peptide #status predicted <ACT>
F;25-246/Product: trypsin V, b-form #status predicted <MAT>
F;25-239/Domain: trypsin homology <TRY>
F;31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted
F;64,108,200/Active site: His, Asp, Ser #status predicted

alignment_scores:
Quality: 237.50 Length: 209
Ratio: 2.220 Gaps: 6
Percent Similarity: 51.196 Percent Identity: 31.100

alignment_block:
US-09-030-606-174 x JQ1472 ..

Align seg 1/1 to: JQ1472 from: 1 to: 246

```
3 TCAGCCGACACTGTTCAGAAAGTGAAGTGCAGAGCTCTACACCATCGG 52
```

```
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 SerAlaAlaHisCysTyrHisProGlnLeuGln.....ValAr 73
53 GCTGGGCGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGG 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 gleuGlyGlnHisAsnIle..TyrGlnIleGlnGlyAlaGlnGlnPheI 89
103 TGGAGCCAGCGCTCTCCGTACGGCACCAGAGTACAAACAGACCCCTTGCTC 152
:::|||||::: |||||:::||||| ||||| ||||| ||||| |||||
89 leAspAlaAlaLysMetIleLeuHisProAspTyrAspLysTrpThrVal 105
153 GCTACGACCTTCATGCTCATCAAGTTGACGAATCCGTGCTGAGTCTGA 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaThrLeuAsnSe 122
203 CACCATCCGAGCATCAGCATTCCTTCGACAGTGCCTACCGCGGGGAAT 252
::: |||||::: |||||::: ||||| ||||| ||||| ||||| |||||
122 rLysValSerThrIleProLeuProGlnTyrCysProThrAlaGlyThrG 139
253 CTGCGCTCGTTCTTGCTGGCTGGGCTGCTGCTGGCGAACGGTGAAGTCA 302
:::|||||::: |||||::: ||||| ||||| ||||| ||||| |||||
139 lucLysLeuValSerGlyTrpGlyValLeuLysPheGly..... 151
303 GTGTGTCTGCCCTCTTCAAGAGAGTCTCTGCCAGTGGCGGGGCTGAC 352
151 ..... 151
353 CCAGAGCTCTGCGTCCAGGAGGAGATG.CCTACCGTGTGCTGAGTGGTGA 401
||| |||:::||||| ||||| ||||| ||||| ||||| |||||
152 .....PheGlnSerProSerValLeuGlnCysLeuA 162
402 ACGTGTCGGTGTCTGTGANGAGTCTGCANTAGCTCTATGACCCGCTG 451
::: |||||::: |||||::: ||||| ||||| ||||| ||||| |||||
162 sPAlaProValLeuSerAspSerValCysHisLysAlaTyrProArgGln 178
452 TACCACCCCANCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAGACTC 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
179 lIeThrAsnAsnMetPheCysLeuGlyPheLeuGlnGlyLysAspSe 195
502 CTGCAACGTGAGAGAGGGGAAAGGGGAGCGGAGCTCAGCGAAGGCT 551
|||||::: |||||::: ||||| ||||| ||||| ||||| |||||
195 rCysGlnTyrAspSerGly..... 201
552 GGAGAAGGGGAGACAGAGACACAGGGCGG.....CATGGCGAG 592
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 .....GlyProValValCysAsnGlyGlu 209
593 ATGCAG.....AGATGCAGAGAC 610
::: |||||::: ||||| ||||| ||||| ||||| ||||| |||||
210 ValGlnGlyIleValSerTrpGlyAsp 218
```

seq_name: pir2:JQ1471

seq_documentation_block:

trypsin (EC 3.4.21.4) V precursor, a-form - rat
C:Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
C;Accession: JQ1471; S23784
R;Kang, J.; Wiegand, U.; Mueller-Hill, B.
Gene 110, 181-187, 1992
A;Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
A;Reference number: JQ1471; MUID:92165057
A;Accession: JQ1471
A:Molecule type: mRNA
A;Residues: 1-246 <KAN>
A;Cross-references: EMBL:X59012; NID:g57412; PID:g57413
A;Experimental source: pancreas
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; protein digestion; serine proteinase; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-24/Domain: activation peptide #status predicted <ACT>
F;25-246/Product: trypsin V, a-form #status predicted <MAT>
F;25-239/Domain: trypsin homology <TRY>

F:31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted
F:64,108,200/Active site: His, Asp, Ser #status predicted

alignment_scores: Quality: 237.50 Length: 209
 Ratio: 2.220 Gaps: 6
Percent Similarity: 51.196 Percent Identity: 31.100

alignment_block:
US-09-030-606-174 x JQ1471 ..

Align seg 1/1 to: JQ1471 from: 1 to: 246

```
3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCTACACCATCGG 52
|||||
61 SerAlaAlaHisCysTyrHisProGlnLeuGln.....ValAr 73
53 GCTGGGCTGCACAGTCTTGAGCCGACCAAGCCAGGAGCCAGATGG 102
|||||
73 GluGlyGluHisAsnIle..TyrGluIleGluGlyAlaGluGlnPheI 89
103 TGGAGGCCACGCTCTCCGTACGGCACCCAGAGTACACAGACCCTGCTC 152
:::
89 IeAspAlaAlaLysMetIleLeuHisProAspTyrAspLysTrpThrVal 105
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTTGCCAGTCTGA 202
|||||
106 AspAsnAspIleMetLeuIleLeuLysSerProAlaThrLeuAsnSe 122
203 CACCATCCGGAGCATCAGCATTTGCTTCGACAGTCCCTACCGGGGAAT 252
:
122 rLysValSerThrIleProLeuProGlnTyrCysProThrAlaGlyThrG 139
253 CTGCTCTGTTTCTGGCTGGGGTCTGCTGGCGAAGGTTAGCTCAGGGGT 302
:::
139 IuCysLeuValSerGlyTyrGlyValLeuLysPheGly..... 151
303 GTGTGTCTGCCCTCTCAAGGAGGTCTCTGCCAGTCCGGGGGCTGAC 352
151 ..... 151
353 CCAGAGCTTGCCTCCAGGACGAATG.CCTACCGTGTGCTGACGTGCTGA 401
|||
152 .....PheGlnSerProSerValLeuGlnCysLeuA 162
402 ACGTGTGGTGTGTCTGANGAGGTTCTGCANTAAAGCTTATGACCCGCTG 451
:::
162 sPaIaProValLeuSerAspSerValCysHisLysAlaTyrProArgGln 178
452 TACCACCCCANCAATGTTCTGCGCCGGCGGAGGGAAGACCAAGAGACTC 501
|||
179 IleThrAsnAsnMetPheCysLeuGlyPheLeuGluGlyGlyLysAspSe 195
502 CTGCAACGTGAGAGAGAGGGAAGGAGGAGGCGAGCTCAGGGAAGGT 551
|||
195 rCysGlnTyrAspSerGly..... 201
552 GGAGAAGGGGAGACAGACACACAGGCGG.....CATGGCGAG 592
|||
202 .....GlnProValValCysAsnGlyGlu 209
593 ATGCAG.....AGATGGAGAGAC 610
:::
210 ValGlnGlyIleValSerTrpGlyAsp 218
```

seq_name: p1r1:TRDGC

seq_documentation_block:
trypsin (EC 3.4.21.4) precursor, cationic - dog
N:Alternate names: cationic trypsinogen
C:Species: Canis lupus familiaris (dog)
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jan-1999

C:Accession: B26273
R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.
Mol. Cell. Biol. 5, 2669-2676, 1985
A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA se
A:Reference number: A26273; MUID:86284628
A:Accession: B26273

A:Molecule type: mRNA
A:Residues: 1-246 <PIN>
A:Cross-References: GB:M11590; NID:g164096; PID:g164097
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-246/Product: trypsin, cationic #status predicted <ENZ>
F:24-239/Domain: trypsin homology <TRY>
F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

alignment_scores: Quality: 236.00 Length: 202
 Ratio: 2.107 Gaps: 7
Percent Similarity: 55.446 Percent Identity: 30.693

alignment_block:
US-09-030-606-174 x TRDGC ..

Align seg 1/1 to: TRDGC from: 1 to: 246

```
3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCTACACCATCGG 52
|||||
60 SerAlaAlaHisCysTyrLysSerArgIleGln.....ValAr 72
53 GCTGGGCTGCACAGTCTTGAGCCGACCAAGACCAGGAGCCAGATGG 102
|||||
72 GluGlyGluTyrAsnIle..AlaValSerGluGlyGluGlnPheI 88
103 TGGAGGCCACGCTCTCCGTACGGCACCCAGAGTACACAGACCCTTGTCTC 152
:::
88 IeAsnAlaAlaLysIleIleArgHisProArgTyrAsnAlaAsnThrIle 104
153 GCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTTGCCAGTCTGA 202
|||||
105 AspAsnAspIleMetLeuIleLysLeuSerSerProAlaThrLeuAsnSe 121
203 CACCATCCGGAGCATCAGCATTTGCTTCGACAGTCCCTACCGGGGAAT 252
:
121 rArgValSerAlaIleAlaLeuProLysSerCysProAlaIaGlyThrG 138
253 CTGCTCTGTTTCTGGCTGGGGTCTGCTGGCGAAGGTGAGCTCAGGGGT 302
:::
138 IuCysLeuIleSerGlyTyrGly..... 145
303 GTGTGTCTGCCCTCTCAAGGAGGTCTCTGCCAGTCCGGGGGCTGAC 352
|||
146 .....AsnThrGlnSerIleGly..... 151
353 CCAGAGCTTGCCTCCAGGACGAATGCTACCGTGTGACGTGCGTGAA 402
|||
152 .....Gln..AsnTyrProAspValLeuGlnCysLeuLy 162
403 CGTGTGCTGCTGTCTGANGAGGTTCTGCANTAAAGCTTATGACCCGCTGT 452
:::
162 sAlaProIleLeuSerAspSerValCysArgAsnAlaTyrProGlyGlnI 179
453 ACCACCCCANCAATGTTCTGCGCCGGCGGGAAGGCAAGACCAAGACTCC 502
|||
503 IeSerSerAsnMetMetCysLeuGlyTyrMetGluGlyGlyLysAspSer 195
503 TGCACAGTGTAGAGAGGGAAGGAGGAGGCAAGGCACTCAGGGAAGGTTG 552
|||
196 CysGlnGlyAspSerGly..... 201
```



```

553 GAGAGGGGGGAGACAGAGACACACAGGGCCG.....CATGGCCAGA 593
      |||||
202 .....GlyProValCysAsnGlyIuL 210
      :|||:
594 TGCAG 598
      :|||
210 euGln 211

```

seq_name: pir4:E40201

seq_documentation_block:

artifact:warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Contents: conceptual translations of ALU-repeat sequences
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CIA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; MUID:92241891
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 233.50 | Length: 103 |
| Ratio: | 3.243 | Gaps: 7 |
| Percent Similarity: | 69.903 | Percent Identity: 56.311 |

alignment_block:

US-09-030-606-174/rev x E40201

Align seg 1/1 to: E40201 from: 1 to: 597

```

1450 TTTT TTTT TGAGACAGAGCTCTTACTCT.. GTTGCCCCAGCTGGAGATATA 1404
||||| ||| ::|||::: |||| | |||||
501 PhePhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyAla 517
||||| ||| ::|||::: |||| | |||||
1403 GTGGTGTGATCTCAACTCA..... CTTCACCTCTGCTC 1369
||| ::|||::: ||| ::|
517 gTPrHisAsnLeuThrAlaAsnPheAlaSerTrpValGlnAlaIle... 532
||||| ||| ::|||::: |||| | |||||
1368 CCATATTCACAGCAATCTCTGCTCAGCCTCCACAGTAGCTGGGATTAC 1319
||||| ||| ::|||::: |||| | |||||
533 ..... LeuSerCysLeuSerLeuProSerSerTrpAspTyr 544
||||| ||| ::|||::: |||| | |||||
1318 AGCGCGCTGCCACCATATCCAGCTTAACCTTTGTATTTTTAGTACAGACAG 1269
||||| ||| ::|||::: |||| | |||||
545 ArgHisAlaProProArgProAlaAsn.PheIlePheLeuValGlnMetG 561
||||| ||| ::|||::: |||| | |||||
1268 GATTTCACCATTTTGGCCAGGCTGCTTGAACCTCCTTACCCTCAAGTCAT 1219
||||| ||| ::|||::: |||| | |||||
561 lYrPheLeuHisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAsp 577
||||| ||| ::|||::: |||| | |||||
1218 CTGCCCTGCCCTGC..... CTCCCAAGTCTGGGATTACAGGCATGAC 1175
||||| ||| ::|||::: |||| | |||||
578 ...Pro.ArgGlnLeuGlnLeuProLysArgTrpAspTyrArgHisGln 593
||||| ||| ::|||::: |||| | |||||

```

seq_name: pir2:S05494

seq_documentation_block:

trypsin (EC 3.4.21.4) IV precursor - rat
 N/Alternate names: 23k protein; trypsinogen IV precursor
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 08-Sep-1997
 C/Accession: S05494
 R/Luetteke, H.; Rausch, U.; Vasiloudes, P.; Scheele, G.A.; Kern, H.F.
 Nucleic Acids Res. 17, 6736, 1989
 A/Title: A fourth trypsinogen (P23) in the rat pancreas induced by CCK.
 A/Reference number: S05494; MUID:89386010
 A/Accession: S05494
 A/Molecule type: mRNA
 A/Residues: 1-247 <LUE>
 A/Cross-references: EMBL:X15679; NID:g56813; PID:g56814
 C/Superfamily: trypsin; trypsin homology
 C/Keywords: calcium binding; hydrolase; protein digestion; serine proteinase; zymogen
 F;1-15/Domain: signal sequence #status predicted <SIG>
 F;16-23/Domain: activation peptide #status predicted <APT>
 F;24-247/Product: trypsin IV #status predicted <MAT>
 F;24-240/Domain: trypsin homology <TRY>
 F;30-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
 F;64,108,201/Active site: His, Asp, Ser #status predicted
 F;76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 233.00 | Length: 173 |
| Ratio: | 2.354 | Gaps: 4 |
| Percent Similarity: | 57.225 | Percent Identity: 32.370 |

alignment_block:

US-09-030-606-174 x S05494 . .

Align seg 1/1 to: S05494 from: 1 to: 247

```

3 TCAGCCGCGACACTGTTTCCAGAAAGTGAAGTGCAGAGCTCCTACACCATCGG 52
  ||||| ||||| ||||| ::::: ::|||
61 SerAlaIaHisCysTyrTyrLysArgLysLeuGln.....ValArg 73
53 GCTGGCGCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGG 102
  ||||| ||::: ::::: || ||::: ||:::
73 gLeuGIygluHisAsnIleHisVal...LeuGIyglYgluGlnPheI 89
103 TGGAGGCGACAGCCTCTCCGTACGCGCACCCAGAGTACACAGACCCCTGTGC 152
  ::::: ||::: ::||| ||||| ||||| ::: |||
89 leAspAlaGluLysIleIleIleArgHisProGluTyrAsnLysAspThrLeu 105
153 GCTACGACACTCATGCTCATCAAGTGGACGGAATCCGTGTCGAGTCTGA 202
  ||||| ::||| ||||| ||||| ::: ::: :::::
106 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaValLeuAsnSe 122
203 CACCATCCGAGACATCAGCATTTGCTTGCAGTGCCTTACCGCGGGGAAT 252
  : ::: ::::: ||::: ::||| ::::: :
122 rGlnValSerThrValSerLeuProArgSerCysAlaSerThrAspAlaG 139
253 CTTGCGCTCGTTTCTGGCGTGGGCTGCTGCGCAACGGTGAGCTCAGGGGT 302
  ::||| ||||| ||||| |||||
139 IncysLeuValSerGlyTrpGly..... 146
303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTTGCCACAGTCGCGGGGCGTGAC 352
  : :
147 .....As 147
353 CCAGAGCTCTGCCGTCCAGGACAGAAATGCCTAACGTGCTGCAGTGCCTGAA 402
  : ||| ::: ||::: ||::: ::||| ||||| :::::
147 nThrValSer..IleGlyGIyLysTyrProAlaLeuLeuGlnCysLeuGI 163
403 CGTGTCCGTGGTGTCTCGANGAGGCTGCANTAAAGCTCTATGACCCGCTGT 452
  ::: ||::: ||::: ::: ||| ||| |||
163 uAlaProValLeuSerAlaSerSerCysLysLysSerTyrProGIyGlnI 180
453 ACCACCCCGCANCATGTCTCTGCGCGCGGGGAGGGCAAGACCAGAGAGACTCC 502

```


This Page Blank (uspto)

OM of: US-09-030-606-174 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:45 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL-frame+_n2p.model -DEV-xlp
-O=/cgn2_1/USPRO_spool/US09030606/runat_24091999_171618_29883/app_query.fasta.1
-DB=SwissProt_37 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-174
Query length: 1459
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

| Sequence | Strd Orig | ZScore | EScore | len | Documentation |
|-------------------------|-----------|--------|---------|-----|---------------------------------|
| SwissProt_37:ALU8_HUMAN | - 359.50 | 518.26 | 4.4e-22 | 591 | P39195 homo sapiens (human). |
| SwissProt_37:ALU8_HUMAN | + 343.50 | 494.57 | 9.3e-21 | 591 | P39195 homo sapiens (human). |
| SwissProt_37:ALU6_HUMAN | - 338.50 | 487.14 | 2.4e-20 | 593 | P39194 homo sapiens (human). |
| SwissProt_37:ALU6_HUMAN | - 324.50 | 466.41 | 3.4e-19 | 593 | P39194 homo sapiens (human). |
| SwissProt_37:ALU5_HUMAN | - 320.00 | 459.86 | 8.0e-19 | 585 | P39192 homo sapiens (human). |
| SwissProt_37:ALU1_HUMAN | - 317.00 | 455.33 | 1.4e-18 | 591 | P39188 homo sapiens (human). |
| SwissProt_37:ALU2_HUMAN | - 314.00 | 450.95 | 2.5e-18 | 587 | P39189 homo sapiens (human). |
| SwissProt_37:ALU7_HUMAN | + 314.00 | 450.86 | 2.5e-18 | 593 | P39194 homo sapiens (human). |
| SwissProt_37:ALU5_HUMAN | + 313.00 | 449.50 | 3.0e-18 | 585 | P39194 homo sapiens (human). |
| SwissProt_37:ALU3_HUMAN | - 312.00 | 447.99 | 3.7e-18 | 587 | P39190 homo sapiens (human). |
| SwissProt_37:ALU1_HUMAN | + 298.00 | 427.20 | 5.2e-17 | 591 | P39188 homo sapiens (human). |
| SwissProt_37:ALU2_HUMAN | + 290.00 | 415.32 | 2.4e-16 | 593 | P39193 homo sapiens (human). |
| SwissProt_37:ALU3_HUMAN | + 286.00 | 409.49 | 5.1e-16 | 587 | P39190 homo sapiens (human). |
| SwissProt_37:ALU4_HUMAN | - 252.00 | 358.91 | 3.3e-13 | 603 | P39191 homo sapiens (human). |
| SwissProt_37:TRYA_RAT | + 237.50 | 345.20 | 4.6e-12 | 246 | P32821 rattus norvegicus (rat). |
| SwissProt_37:TRYB_RAT | + 237.50 | 345.20 | 4.6e-12 | 246 | P32822 rattus norvegicus (rat). |
| SwissProt_37:TRY1_CANFA | + 236.00 | 342.98 | 6.2e-12 | 246 | P06871 canis familiaris (dog). |
| SwissProt_37:TRY4_RAT | + 233.00 | 338.50 | 1.1e-11 | 247 | P12788 rattus norvegicus (rat). |
| SwissProt_37:TRY1_BOVIN | + 226.50 | 329.02 | 3.7e-11 | 243 | P00760 bos taurus (bovine). |
| SwissProt_37:TRY3_RAT | + 224.50 | 325.91 | 5.5e-11 | 247 | P08426 rattus norvegicus (rat). |
| SwissProt_37:SCCE_HUMAN | + 222.00 | 322.01 | 8.8e-11 | 253 | P49862 homo sapiens (human). |
| SwissProt_37:TRY2_XENLA | + 221.50 | 321.58 | 9.7e-11 | 244 | P70059 xenopus laevis (afrod). |
| SwissProt_37:TRY1_HUMAN | + 221.50 | 321.47 | 9.7e-11 | 247 | P07477 homo sapiens (human). |
| SwissProt_37:TRY3_SALSA | + 220.50 | 320.31 | 1.2e-10 | 238 | P35033 salmo salar (atlantic). |
| SwissProt_37:TRY1_CHICK | + 220.50 | 319.96 | 1.2e-10 | 248 | P09062 gallus gallus (chicken). |
| SwissProt_37:TRY2_CHICK | + 220.50 | 319.96 | 1.2e-10 | 248 | P09062 gallus gallus (chicken). |
| SwissProt_37:TRY3_CHICK | + 219.00 | 317.74 | 1.6e-10 | 248 | P09062 gallus gallus (chicken). |
| SwissProt_37:TRY1_SALSA | + 218.50 | 309.80 | 4.4e-10 | 242 | P35032 salmo salar (atlantic). |
| SwissProt_37:TRY1_XENLA | + 214.00 | 302.64 | 4.4e-10 | 603 | P39191 homo sapiens (human). |
| SwissProt_37:TRY1_XENLA | + 211.50 | 306.81 | 6.5e-10 | 243 | P19799 xenopus laevis (afrod). |
| SwissProt_37:TRY4_HUMAN | + 211.50 | 304.87 | 6.6e-10 | 304 | P35030 homo sapiens (human). |
| SwissProt_37:TRY1_RAT | + 208.50 | 302.26 | 1.1e-09 | 246 | P00762 rattus norvegicus (rat). |
| SwissProt_37:TRYP_MOUSE | + 208.00 | 301.52 | 1.3e-09 | 246 | P07146 mus musculus (mouse). |
| SwissProt_37:TRYP_PIG | + 207.50 | 301.32 | 1.4e-09 | 231 | P00761 sus scrofa (pig). |
| SwissProt_37:TRYP_SQUAC | + 207.00 | 300.66 | 1.5e-09 | 229 | P00764 squalus acanthias (sp). |
| SwissProt_37:TRY2_HUMAN | + 206.50 | 299.26 | 1.7e-09 | 247 | P07478 homo sapiens (human). |
| SwissProt_37:KLB_RAT | + 206.00 | 298.11 | 1.9e-09 | 259 | P36376 rattus norvegicus (rat). |
| SwissProt_37:TRY2_RAT | + 202.00 | 292.63 | 3.9e-09 | 246 | P00763 rattus norvegicus (rat). |
| SwissProt_37:KKA_MOUSE | + 202.00 | 292.12 | 4.0e-09 | 261 | P15946 mus musculus (mouse). |
| SwissProt_37:TRY2_BOVIN | + 197.00 | 285.19 | 1.0e-08 | 247 | P29463 bos taurus (bovine). |

SwissProt_37:TRYX_GADMO + 196.50 284.67 1.1e-08 241 ! Q91041 gadus morhua (atla
SwissProt_37:TRY1_GADMO + 195.50 283.19 1.3e-08 241 ! P16049 gadus morhua (atla
SwissProt_37:TRY2_CANFA + 195.00 282.23 1.5e-08 247 ! P06872 canis familiaris (

seq_name: SwissProt_37:ALU8_HUMAN

seq_documentation_block:

ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE !!! ALU SUBFAMILY SX WARNING ENTRY !!!
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA OQUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -1- CAUTION: ALU REPEITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC DNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CONAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U14574; -; NOT_ANNOTATED_CDS.

KW HYPOTHETICAL PROTEIN.
 FT DOMAIN 1 96 FRAME 1.
 FT DOMAIN 100 195 FRAME 2.
 FT DOMAIN 199 294 FRAME 3.
 FT DOMAIN 298 393 FRAME 4.
 FT DOMAIN 397 492 FRAME 5.
 FT DOMAIN 496 591 FRAME 6.
 SQ SEQUENCE 591 AA; 64395 MW; ABF69224 CRC32;

alignment_scores:
 Quality: 359.50 Length: 96
 Ratio: 4.609 Gaps: 1
 Percent Similarity: 81.250 Percent Identity: 73.958

alignment_block:

US-09-030-606-174/rev x ALU8_HUMAN ..

Align seg 1/1 to: ALU8_HUMAN from: 1 to: 591

1447 TTTTGGAGACAGAGTCTTACTCTGTGCCCCAGCTGGAGTATAGTGCTG 1398
 496 PhepNegluthrGluSerArgSerValAlaGlnAlaGlyValGlnTrpPar 512
 1397 TGATCTCACTCACTCACTCACTGCTGCCCATATTCAGCAATTCCTCT 1348
 512 GAspLeuGlySerLeuGlnProProProGlyPheLysArgPheSerC 529
 1347 GCCTCAGCCTCCCAAGTAGCTGGATTACAGGCGCTGCCACCATATCCA 1298
 529 yslEulerProSerSerTrpAspIyAlaArgAlaProProArgPro 545
 1297 GCTACTTTGATTTTGTAGTACAGACAGAGATTTCACCATTTGGCCAGG 1248
 546 AlaAspPheCysIlePheSerArgAspGlyValSerProCysTrpProG1 562
 1247 CTGCTCTGAACCTCTTACTCAAGTAGTCTGCTGCTGCCCTGCCCAAA 1198
 562 yTrpSerArgThrProAspLeuArg**Ser.....SerSerGlnS 576
 1197 GTGCTGGATTACAGGATGAGCCACCCGCCCCAGCCT 1160
 576 eraLaGlyIleThrGlyValSerHisArgAlaArgPro 588

seq_name: SwissProt_37:ALU8_HUMAN

seq_documentation_block:

ID ALU8_HUMAN STANDARD; PRT; 591 AA.

AC P39195;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE 1111 ALU SUBFAMILY SX WARNING ENTRY 1111
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95021758.
 RA CLAVERIE J.-M., MAKALOWSKI W.;
 RT "Alu alert."
 RL NATURE 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE; 92241891.
 RA CLAVERIE J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";
 RL GENOMICS 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE; 88333009.

RA QUENTIN Y.;
 RT "The Alu family developed through successive waves of fixation
 closely connected with primate lineage history."
 RL J. MOL. EVOL. 27:194-202(1988).

RP [4]
 RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE; 91178815.

RA JURKA J., MILOSAVLJEVIC A.;
 RT "Reconstruction and analysis of human Alu genes."
 RL J. MOL. EVOL. 32:105-121(1991).

CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
 CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
 CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
 CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
 CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
 CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.

CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.

CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U14574; -; NOT_ANNOTATED_CDS.

KW HYPOTHETICAL PROTEIN.
 FT DOMAIN 1 96 FRAME 1.
 FT DOMAIN 100 195 FRAME 2.
 FT DOMAIN 199 294 FRAME 3.
 FT DOMAIN 298 393 FRAME 4.
 FT DOMAIN 397 492 FRAME 5.
 FT DOMAIN 496 591 FRAME 6.
 SQ SEQUENCE 591 AA; 64395 MW; ABF69224 CRC32;

alignment_scores:
 Quality: 343.50 Length: 97
 Ratio: 4.294 Gaps: 1
 Percent Similarity: 82.474 Percent Identity: 72.165

alignment_block:

US-09-030-606-174 x ALU8_HUMAN ..

Align seg 1/1 to: ALU8_HUMAN from: 1 to: 591

1157 AAGAGGCTGGCAGGCTGCTCATGCTGTATCCACGACACTTTGGAGG 1206

[illegible]

```

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC
CC -! CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; U14573; -; NOT_ANNOTATED_CDS.
CC DR HYPOTHETICAL PROTEIN.
CC KW
CC FT DOMAIN 1 97 FRAME 1.
CC FT DOMAIN 101 196 FRAME 2.
CC FT DOMAIN 200 295 FRAME 3.
CC FT DOMAIN 299 395 FRAME 4.
CC FT DOMAIN 399 494 FRAME 5.
CC FT DOMAIN 498 593 FRAME 6.
CC SQ SEQUENCE 593 AA; 64417 MW; 56C24239 CRC32;

```



```

AC P39193;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE 1111 ALU SUBFAMILY SP WARNING ENTRY 1111
OS HOMO SAPIENS (HUMAN).
OC EURAZOYA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 9117815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -I- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -I- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -I- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -I- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -I- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14572; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 1 97 FRAME 1.

```

```

FT DOMAIN 101 196 FRAME 2.
FT DOMAIN 200 295 FRAME 3.
FT DOMAIN 299 395 FRAME 4.
FT DOMAIN 399 494 FRAME 5.
FT DOMAIN 498 593 FRAME 6.
SQ SEQUENCE 593 AA; 64603 MW; 74982C2D CRC32;

alignment_scores:
    Quality: 324.50      Length: 97
    Ratio: 4.214        Gaps: 1
Percent Similarity: 79.381      Percent Identity: 69.072

alignment_block:
US-09-030-606-174/rev x ALU6_HUMAN ..

Align seg 1/1 to: ALU6_HUMAN from: 1 to: 593

1447 TTTTGGAGACAGATCTTACTCT...GTGCCCCAGCTGGAGTATAGTG 1401
||||| ::|||::| ||||| ||||| ||
299 PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValGlnTr 315
1400 GTGTGATCTCAACTCACTTCAACCTGTGCCTCCCATATTCAGCAATTCT 1351
| |||||::||| ||||| ||||| |||||::|||
315 pArgAspLeuGlySerProGlnProProProGlyPheGlySerArgPheS 332
1350 CCTGCCTCAGCCCTCCCAAGTAGCTGGGATTACAGGCGCGCTGCCACCATAT 1301
||||| ||||| ||||| ||||| ||||| |||||
332 eRcYsLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArg 348
1300 CCAGCTACTTTTGTATTTTGTAGTACAGACAGGATTTACCATTTTGCGC 1251
||||| ||||| ||||| ||||| ||||| |||||
349 ProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpSe 365
1250 AGGCTGGTCTTGAACTCCTTACCTCAAGTGATCTGCCTGCTCGC.CTCC 1202
||||| ||||| ||||| ||||| ||||| |||||
365 rGlyTrpSerArgThrProAspLeuArg**SerAlaArgLeuGlyLeuP 382

1201 CAAAGTGTGGGATTACAGGATGAGCCACCCCTGCCAGCC 1161
||||| ||||| ||||| ||||| ||||| |||||
382 rOlYsCysTrpAspTyrArgGluProProArgProAla 395

seq_name: SwissProt_37:ALU5_HUMAN

seq_documentation_block:
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE |||| ALU SUBFAMILY SC WARNING ENTRY ||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";

```



```

RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE. 91178815.
RA JURKA J., MIOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14571; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 1 95 FRAME 1.
FT DOMAIN 99 193 FRAME 2.
FT DOMAIN 197 291 FRAME 3.
FT DOMAIN 295 389 FRAME 4.
FT DOMAIN 393 487 FRAME 5.
FT DOMAIN 491 585 FRAME 6.
SQ SEQUENCE 585 AA; 63957 MW; 5C794090 CRC32;

alignment_scores:
    Quality: 320.00      Length: 97
    Ratio: 4.000         Gaps: 2
    Percent Similarity: 82.474      Percent Identity: 71.134

alignment_block:
US-09-030-606-174/rev x ALU5_HUMAN ..

Align seg 1/1 to: ALU5_HUMAN from: 1 to: 585

1447 TTTTGTGAGACAGAGTCTTACTCTGTGCCCCAGCTGGAGTATAGTGTG 1398
      ||||||||||||||| ||||||| :::||||| |||||:
491 PhepneuThruGluserArgSerValAla.ArgLeuGlucysSerGly 507
      |||||||:||||| ||||| ||||| |||||:|||||
1397 TGAATCTCAACTCACTTCAACCTCTGCCTCCCATATATCAAGCAATTCTCCT 1348
      ::|||||:||||| ||||| ||||| |||||:|||||
507 laIleSerAlaHisCysAsnLeuArgLeuProGlySerSerAspSerPro 523

1347 GCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGGCGCTGCCACCATATATCCA 1298

```

```

524 AlaSerAlaSerArgValAlaGlyThrThrGlyAlaAlaRghHisAlaGl 540
1297 GCTAACTTTGTATTTTGTAGTACAGACAGAGATTTCACATTTGGCCAG 1248
||||| ||||| ||||| ||||| ||||| |||||
540 nleutIephelvalPheleuValgluThrglYphenishisValglylna 557
1247 CTGCTCTTGAACTCCTTAAGTCAAGTGAATCGCTGCCTGCCCTCCCAA 1198
|||||:::||||| |||||:::||| ::||| |||
557 spgLyLeuAspleuLeuThrSer**SerAlaArgleugly.LeuProly 573
1197 GTGCTGGGATATACAGGACATGACCACCCCTGCCAGCC 1161
||||| ||||| ||||| ||||| |||||
573 scYstPrasptYrArgArgGluProProArgProAla 585
seq_name: SwissProt_37:ALU1_HUMAN
seq_documentation_block:
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE !!! ALU SUBFAMILY J WARNING ENTRY !!!
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOMSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -I- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -I- ISOLATED 'x' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -I- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLYMERIZATION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -I- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

```


CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; U14567; -; NOT_ANNOTATED_CDS.

CC KW HYPOTHETICAL PROTEIN.

FT DOMAIN 1 96 FRAME 1.
 FT DOMAIN 100 195 FRAME 2.
 FT DOMAIN 199 294 FRAME 3.
 FT DOMAIN 298 393 FRAME 4.
 FT DOMAIN 397 492 FRAME 5.
 FT DOMAIN 496 591 FRAME 6.
 SQ SEQUENCE 591 AA; 63790 MW; AAD89112 CRC32;

alignment_scores:

Quality: 317.00 Length: 96
 Ratio: 4.284 Gaps: 0
 Percent Similarity: 77.083 Percent Identity: 69.792

alignment_block:

us-09-030-606-174/rev x ALU1_HUMAN ..

Align seg 1/1 to: ALU1_HUMAN from: 1 to: 591

1447 TTTTGTGAGACAGAGTCTTACTGTGCCCCAGCTGAGTATAGTG 1398
 |||||
 496 PhepHeGIuThrGIySerArGSeValAlaGlnAlaGlyValGlnTrpAr 512
 1397 TGATCTCACTCACTCAACCTCTGCCCTCCCATATTCAGCAATTCTCT 1348
 ||| ::|||
 512 GAspHisGIySerLeuGlnProArGProGIyLeuLysArGSeSer 529
 1347 GCCTCAGCCTCCCAAGTAGCTGGGATTACAGCGCCCTGCCACCATATCCA 1298
 |||||
 529 ySleuSerLeuProSeSerTrpAspTyArGArGAlaProProArGPro 545
 1297 GCTAACTTTGTATTTTAGTACAGACAGATTCACCATTTGGCCAGG 1248
 |||||
 546 AlaAspHeCySilePheCySArGAspGIyValSerLeuCySProGI 562
 1247 CTGGTCTTGACTCGCTTACCTCAAGTATGATCGCTGCCTGC. CTCCCA 1199
 |||||
 562 yTrpSerArGThrProGIyLeuLys***SerSerArGLeuGIyLeuProL 579
 1198 AGTGTGGGATTACAGCATGAGCCACCCCTGCCAGCC 1161
 |||||
 579 yScySTrpAspTyArGArGAlaProProArGProAla 591

seq_name: SwissProt_37:ALU2_HUMAN

seq_documentation_block:

ID ALU2_HUMAN STANDARD: PRT; 587 AA.
 AC P39189;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE |||| ALU SUBFAMILY SB WARNING ENTRY ||||
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95021758.
 RA CLAVERIE J.-M., MAKALOWSKI W.;
 RT "Alu alert.";
 RL NATURE 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE; 92241891.
 RA CLAVERIE J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 RT potentially misleading protein sequences.";
 RL GENOMICS 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE; 88333009.
 RA OUBENIN Y.;
 RT "The Alu family developed through successive waves of fixation
 RT closely connected with primate lineage history.";
 RL J. MOL. EVOL. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE; 91178815.
 RA JURKA J., MILOSAVLJEVIC A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. MOL. EVOL. 32:105-121(1991).
 CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
 CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
 CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
 CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
 CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
 CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U14568; -; NOT_ANNOTATED_CDS.

KW HYPOTHETICAL PROTEIN.

FT DOMAIN 1 96 FRAME 1.
 FT DOMAIN 100 194 FRAME 2.
 FT DOMAIN 198 292 FRAME 3.
 FT DOMAIN 296 391 FRAME 4.
 FT DOMAIN 395 489 FRAME 5.
 FT DOMAIN 493 587 FRAME 6.
 SQ SEQUENCE 587 AA; 63703 MW; 844F0DCE CRC32;

alignment_scores:

Quality: 314.00 Length: 95
Ratio: 4.486 Gaps: 1
Percent Similarity: 73.684 Percent Identity: 70.526

alignment_block:

US-09-030-606-174/rev x ALU2_HUMAN ..

Align seg 1/1 to: ALU2_HUMAN from: 1 to: 587

1447 TTTTGTGAGACAGAGTCTACTCTGTGCCCCAGCTGGAGTATAGTGTG 1398
|||||
493 phepегluthrgluserarqservalaiaaglinalaglyvalglntrpar 509
1397 TGATCTCACTCACTCAACCTCTGCCTCCCATATTCACGCAATTCCT 1348
|||||
509 gaspleuglyserleuglnalapropropoglyphethrpropheserc 526
1347 GCCTCAGCCTCCCAAGTAGCTGGATTACAGCGCGCTGCCACCATATCCA 1298
|||||
526 ysleuserleuprosersertrpasptryargrproproargpro 542
1297 GCTAAGTTTGTATTTTGTAGTACAGACAGATTTCACCAATTTGCCAGG 1248
|||||
543 AlaasnphecysilepheserArqspclyValserProcys***Progl 559
1247 CTGCTCTGACTCCTTACTCAAGTATCTGCCTGCCTGCCCTCCCAA 1198
|||||
559 ytrpserarqserproaspLeu.Valileargproproargproprolys 575
1197 GTGCTGGATTACAGGCATGAGCCACCTGCC 1165
|||||
576 ValleuglyleuglnAla**AlatrhAlaPro 586

seq_name: SwissProt_37:ALU7_HUMAN

seq_documentation_block:

ID ALU7_HUMAN STANDARD; PRT; 593 AA.

AC P39194;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE |||| ALU SUBFAMILY SQ WARNING ENTRY ||||

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 95021758.

RA CLAVERIE J.-M., MAKALOWSKI W.;

RT "Alu alert.";

RL NATURE 371:752-752(1994).

RN [2]

RP CONCEPT.

RX MEDLINE; 92241891.

RA CLAVERIE J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";

RL GENOMICS 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE; 88333009.

RA QUENTIN Y.;

RT "The Alu family developed through successive waves of fixation

RT closely connected with primate lineage history.";

RL J. MOL. EVOL. 27:194-202(1988).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE; 91178815.

RA JURKA J., MILOSAVLJEVIC A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. MOL. EVOL. 32:105-121(1991).

CC -!- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS

CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -!- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U14573; -; NOT_ANNOTATED_CDS.

KW HYPOTHETICAL PROTEIN.

FT DOMAIN

FT DOMAIN 1 97 FRAME 1.

FT DOMAIN 101 196 FRAME 2.

FT DOMAIN 200 295 FRAME 3.

FT DOMAIN 299 395 FRAME 4.

FT DOMAIN 399 494 FRAME 5.

FT DOMAIN 498 593 FRAME 6.

SEQUENCE 593 AA; 64417 MW; 56C24239 CRC32;

alignment_scores:

Quality: 314.00 Length: 97
Ratio: 4.026 Gaps: 0
Percent Similarity: 80.412 Percent Identity: 70.103

alignment_block:

US-09-030-606-174 x ALU7_HUMAN ..

Align seg 1/1 to: ALU7_HUMAN from: 1 to: 593

1161 GGCTGGGACGGGTGCTCATGCTGTAATCCACGACTTTGGAGGC.GA 1209
|||||
1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGl 17
1210 GGCAGGAGATCACTTGAGGTAAAGAGTCAAGACGACCTGGCCAAAT 1259
|||||
17 uAlaGlyGlySerProGluValArgSerSerArgProAlaTrpProThrT 34
1260 GGTGAATCCTGTCTGTACTAAATAACAAAGTACCTGGATATGTTG 1309
|||
34 rp***AsnProValSerThrLysAsnThrLysIleSerArgAlaTrpTrp 50
1310 CAGCGCCTGTAATCCACGACTTGGAGGCTGAGCAGACGAGATTCCT 1359
:::|||||
51 ArgAlaProValIleProAlaThrArgGluAlaGluAlaGlyGlySerIle 67
1360 TGAATATGGGAGGACGAGAGGTGAAGTGAAGTGAAGATCAACACCACTATACT 1409


```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RL potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RL closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14569; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 1 96 FRAME 1.
FT DOMAIN 100 194 FRAME 2.
FT DOMAIN 198 292 FRAME 3.
FT DOMAIN 296 391 FRAME 4.
FT DOMAIN 395 489 FRAME 5.
FT DOMAIN 493 587 FRAME 6.
SQ SEQUENCE 587 AA; 63573 MW; CA32BFF1 CRC32;

```

```

alignment_scores:
    Quality: 312.00      Length: 95
    Ratio: 4.457        Gaps: 1
    Percent Similarity: 73.684      Percent Identity: 70.526
alignment_block:
US-09-030-606-174/rev x ALU3_HUMAN ..
Align seg 1/1 to: ALU3_HUMAN from: 1 to: 587
1447 TTTTGTGAGACAGAGTCTTACTCTGTGCCCCAGCTGGAGTATAGTGTG 1398
|||||
493 PhePheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpPar 509
1397 TGATCTCAACTCACTTCAACCTCTGCTCCCATATTCAGCAATTCCT 1348
|||||
509 GasPheGlySerLeuGlnAlaProProGlyPheThrProPheSerC 526
1347 GCCTCAGCCTCCCAAGTAGCTGGGATTACAGCGCGCTGCCACCATATCCA 1298
|||||
526 YsLeuSerLeuProSerSerTrpAspTyrArgArgProProLeuArgPro 542
1297 GCTACTTTGTGATTTTGTAGTACAGACAGAGATTTCACCATTTGGCCAGC 1248
|||||
543 AlaAsnPhcCysIlePheSerArgAspGlyValSerProPhe**ProGI 559
1247 CTGCTCTGAACCTCTTACTCAAGTGATCTGCTGCTGCTGCCCTCCCAAA 1198
|||||
559 YTrpSerArgSerProAspLeu.ValIleArgProProArgProProLys 575
1197 GTGCTGGATTACAGGACATGAGCCACCCCTGCC 1165
|||||
576 ValLeuGlyLeuGlnAla**AlaThrAlaPro 586
seq_name: SwissProt_37:ALU1_HUMAN
seq_documentation_block:
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE |||| ALU SUBFAMILY J WARNING ENTRY ||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RL potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RL closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS

```


CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U14567; -; NOT_ANNOTATED_CDS.
CC KW HYPOTHETICAL PROTEIN.
CC FT DOMAIN 1 96 FRAME 1.
CC FT DOMAIN 100 195 FRAME 2.
CC FT DOMAIN 199 294 FRAME 3.
CC FT DOMAIN 298 393 FRAME 4.
CC FT DOMAIN 397 492 FRAME 5.
CC FT DOMAIN 496 591 FRAME 6.
CC SQ SEQUENCE 591 AA; 63790 MW; AAD89112 CRC32;

alignment_scores: Quality: 298.00 Length: 96
 Ratio: 3.921 Gaps: 0
Percent Similarity: 79.167 Percent Identity: 68.750

alignment_block:
US-09-030-606-174 x ALU1_HUMAN ..

Align seg 1/1 to: ALU1_HUMAN from: 1 to: 591

1161 GGCTGGGAGGCTGCTCATGCTGCTATCCAGACACTTGGAGGC.GA 1209
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 GlyArgAlaArgTyrPheThrProValIleProAlaLeuTyrGluAlaG1 17

1210 GGCAGGCAGATCACTGAGTAAGAGAGTCAAGACAGCCCTGGCCAAAT 1259
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uAlaGlyGlySerLeuGluProArgSerSerArgProAlaTyrPalatThr* 34

1260 GGTGAATCCTGTCTGTACTAAATAACAAAGTTAGCTGATATGGTGG 1309
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 *****AsnProValSerThrLysAsnThrLysIleSerArgAlaTyrTrp 50

1310 CAGCGCCTGTATCCAGACTACTTGAGAGCTGAGGAGAGAGAAATTGCT 1359
 :::||||| :::||||| :::||||| :::||||| :::||||| :::|||||
51 ArgAlaProValValProAlaThrArgGluAlaGluAlaGlyGlySerLe 67

1360 TGAATATGGAGGAGGAGAGGTTGAAGTGAGTTGAGATCACACCACTATACT 1409

seq_name: swissprot_37:ALU6_HUMAN
seq_documentation_block:
ID ALU6_HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE |||| ALU SUBFAMILY SP WARNING ENTRY ||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert."
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences."
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history."
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----


```
||||| ::| ||||| |||||::||| |||||
265 roglyglyGlyAlaCysSerGluProArgSerArgHisCysThrProAla 281
1416 GGGCAACAGAGTAACTCTGTCTCAAAAAA 1448
||||| ||||| ||||| |||||
282 TrpAlaThrGluArgAspSerValSerLysLys 292
```

seq_name: SwissProt_37:ALU4_HUMAN

seq_documentation_block:

```
ID ALU4_HUMAN STANDARD: PRT: 603 AA.
AC P39191;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ||||| ALU SUBFAMILY SB2 WARNING ENTRY |||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U14570; -; NOT_ANNOTATED_CDS.

KW HYPOTHETICAL PROTEIN.

| | | | |
|-------------|---------|-----------|-----------------|
| FT DOMAIN | 1 | 98 | FRAME 1. |
| FT DOMAIN | 102 | 199 | FRAME 2. |
| FT DOMAIN | 202 | 300 | FRAME 3. |
| FT DOMAIN | 304 | 401 | FRAME 4. |
| FT DOMAIN | 405 | 502 | FRAME 5. |
| FT DOMAIN | 506 | 603 | FRAME 6. |
| SQ SEQUENCE | 603 AA; | 65272 MW; | A6C07F68 CRC32; |

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 252.00 | Length: | 99 |
| Ratio: | 3.150 | Gaps: | 2 |
| Percent Similarity: | 80.808 | Percent Identity: | 63.636 |

alignment_block:

US-09-030-606-174/rev x ALU4_HUMAN ..

Align seg 1/1 to: ALU4_HUMAN from: 1 to: 603

```
1447 TTTTGTGACAGAGTCTTACTCTGTGCCCCAGCTGGAGTATA..... 1404
||||| ||||| ||||| ||||| |||||
506 PhePheGluThrGluSerArgSerValAlaGlnAlaGlyLeuArgThrAl 522
1403 GTGTGTGATCTCAACTCACTTCAACCTGTGCTCCCATATTCAGCAA 1355
||||| ||||| ||||| ||||| |||||
522 aValAlaGlnSerArgLeuThrAlaSerSerAlaSerArgValHisAlaI 539
1354 TTCTCTGCTCAGCCTCCCAAGTAGCTGGATTACAGCGCCTGCCACC 1305
||||| ||||| ||||| ||||| |||||
539 leuLeuProGlnProGln**LeuGlyLeuGlnAlaProAlaThr 555
1304 ATATCAGCTA.ACTTTGTATTTTAGTACAGACAGATTTCCACCATTT 1256
::: ||||| ||||| ||||| ||||| |||||
556 AlaProGly**PhePheValPheLeuValGluThrGlyPheHisLeuVa 572
1255 TGCCAGGCTGCTTGAAGCTCTTACCTCAAGTGAATCTGCTGCC 1206
::: ||||| ||||| ||||| ||||| |||||
572 IserGlnAspGlyLeuAspLeuThrSer**SerThrArgLeuGly. 588
1205 CTCCCAAGTCTGGATTACAGCATGAGCCACCTGCCAGCC 1161
||||| ||||| ||||| ||||| |||||
589 leuProLysCysTrpAspTyrArgArgGluProProArgProAla 603
```


OM of: US-09-030-606-174 to: SPTREMBL_10:* out_format : pfs
Date: Sep 25, 1999 10:35 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL-frame+₁np.model -DEV-xlp
-Q-/cgn2_1/USPTO_spool/US09030606/runat_24091999_171617_29869/app_query.fasta.1
-DB-SPTREMBL_10 -QFMT-fastan -SUFFIX-rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-biosum62
-TRANS-human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-174
Query length: 1459
Database: SPTREMBL_10:*
Database sequences: 201082
Database length: 61543640
Search time (sec): 260.540000

| Sequence | Strd | Orig | ZScore | Escore | Len | Documentation |
|------------------------|------|--------|--------|---------|------|------------------------------------|
| sp_rodent:Q9Z0M1 | + | 477.50 | 835.71 | 4.7e-39 | 255 | Q9Z0M1 mus musculus (mouse). en |
| sp_human:P78525 | - | 310.00 | 529.12 | 2.1e-22 | 666 | P78525 homo sapiens (human). my |
| sp_human:O14608 | - | 280.00 | 471.61 | 2.1e-19 | 1079 | O14608 homo sapiens (human). v |
| sp_human:Q99770 | - | 270.50 | 471.80 | 1.6e-18 | 139 | Q99770 homo sapiens (human). hy |
| sp_human:O60448 | - | 268.00 | 459.05 | 3.0e-18 | 375 | O60448 homo sapiens (human). ne |
| sp_human:Q15662 | - | 257.00 | 439.60 | 3.7e-17 | 368 | Q15662 homo sapiens (human). tr |
| sp_invertebrate:Q26195 | - | 247.00 | 426.87 | 3.5e-16 | 200 | Q26195 plasmodium vivax. pval |
| sp_human:Q07826 | - | 246.50 | 431.77 | 3.7e-16 | 100 | Q07826 homo sapiens (human). x |
| sp_mhc:Q29976 | + | 242.00 | 418.13 | 1.1e-15 | 196 | Q29976 homo sapiens (human). ma |
| sp_vertibrate:Q92099 | + | 221.00 | 378.93 | 1.4e-13 | 242 | Q92099 paranthothenia magellani |
| sp_vertibrate:Q93265 | + | 220.00 | 376.88 | 1.7e-13 | 250 | Q93265 pleuronectes americanus |
| sp_vertibrate:Q91515 | + | 219.50 | 376.43 | 1.9e-13 | 237 | Q91515 fuigu rubripes (japanese |
| sp_vertibrate:Q92046 | + | 219.00 | 375.13 | 2.1e-13 | 249 | Q92046 dissostichus mawsoni. pr |
| sp_vertibrate:Q93266 | + | 213.50 | 365.56 | 7.5e-13 | 242 | Q93266 pleuronectes americanus |
| sp_vertibrate:Q93594 | + | 212.50 | 366.35 | 9.3e-13 | 178 | Q93594 dicentrarchus labrax (eu |
| sp_human:Q15665 | + | 211.50 | 361.43 | 1.2e-12 | 259 | Q15665 homo sapiens (human). tr |
| sp_rodent:Q63274 | + | 211.00 | 361.35 | 1.3e-12 | 235 | Q63274 rattus norvegicus (rat). se |
| sp_human:Q75837 | + | 211.00 | 359.83 | 1.4e-12 | 282 | Q75837 homo sapiens (human). se |
| sp_rodent:Q921R9 | + | 210.00 | 359.19 | 1.7e-12 | 246 | Q921R9 mus musculus (mouse). tr |
| sp_human:Q15946 | + | 202.00 | 345.75 | 1.0e-11 | 223 | Q15946 homo sapiens (human). gl |
| sp_vertibrate:Q42160 | + | 201.50 | 344.07 | 1.2e-11 | 245 | Q42160 petromyzon marinus (sea |
| sp_vertibrate:Q42608 | + | 201.50 | 344.00 | 1.2e-11 | 247 | Q42608 petromyzon marinus (sea |
| sp_rodent:Q63275 | + | 200.50 | 342.50 | 1.5e-11 | 239 | Q63275 rattus norvegicus (rat). se |
| sp_vertibrate:Q42158 | + | 200.50 | 342.22 | 1.5e-11 | 247 | Q42158 petromyzon marinus (sea |
| sp_rodent:Q03955 | + | 199.50 | 340.34 | 1.9e-11 | 250 | Q03955 pracomys natalensis (afri |
| sp_vertibrate:Q91036 | + | 196.50 | 336.10 | 3.7e-11 | 219 | Q91036 gadus morhua (atlantic d |
| sp_vertibrate:Q42159 | + | 195.50 | 333.41 | 4.6e-11 | 244 | Q42159 petromyzon marinus (sea |
| sp_mammal:Q29474 | + | 191.00 | 324.83 | 1.3e-10 | 261 | Q29474 canis familiaris (dog). v |
| sp_human:Q13675 | - | 189.50 | 316.74 | 1.9e-10 | 499 | Q13675 homo sapiens (human). al |
| sp_human:Q15096 | - | 187.50 | 319.50 | 2.9e-10 | 234 | Q15096 homo sapiens (human). pr |
| sp_rodent:Q88309 | + | 186.50 | 316.81 | 3.6e-10 | 261 | Q88309 mus musculus (mouse). pr |
| sp_invertebrate:Q18599 | + | 186.00 | 316.08 | 4.1e-10 | 256 | Q18599 drosophila virilis (fr |
| sp_rodent:Q61855 | + | 185.50 | 315.02 | 4.6e-10 | 261 | Q61855 mus musculus (mouse). tr |
| sp_human:Q13629 | + | 184.50 | 319.87 | 7.1e-10 | 118 | Q13629 homo sapiens (human). hy |
| sp_rodent:Q54854 | + | 173.50 | 293.96 | 7.1e-09 | 251 | Q54854 rattus norvegicus (rat). se |
| sp_rodent:Q61388 | + | 172.50 | 292.14 | 8.9e-09 | 252 | Q61388 mus musculus (mouse). cy |
| sp_invertebrate:Q27761 | + | 172.50 | 291.69 | 8.9e-09 | 266 | Q27761 penaeus vannamei (pend |
| sp_invertebrate:Q76520 | + | 170.50 | 288.51 | 1.4e-08 | 254 | Q76520 stomoxys calcitrans (s |
| sp_invertebrate:Q62561 | + | 169.50 | 286.41 | 1.8e-08 | 264 | Q62561 penaeus vannamei (pend |
| sp_invertebrate:Q01309 | + | 169.00 | 286.21 | 2.0e-08 | 243 | Q01309 botryllus schlosseri. |
| sp_human:Q95809 | - | 168.00 | 296.69 | 2.2e-08 | 56 | Q95809 homo sapiens (human). bet |
| sp_rodent:Q88301 | + | 165.50 | 279.87 | 4.4e-08 | 246 | Q88301 mus musculus (mouse). se |

seq_name: sp_rodent:Q9Z0M1
seq_documentation_block:
ID Q9Z0M1 PRELIMINARY; PRT; 255 AA.
AC Q9Z0M1;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS-WEBSTER;
RA Simmer J.;
RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF019979; AAC98894.1; -
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.
SQ SEQUENCE 255 AA; 27488 MW; 6E711616 CRC32;

alignment_scores:
Quality: 477.50 Length: 173
Ratio: 3.820 Gaps: 3
Percent Similarity: 72.254 Percent Identity: 55.491

alignment_block:
US-09-030-606-174 x Q9Z0M1 ..

Align seg 1/1 to: Q9Z0M1 from: 1 to: 255

```
3 TCAGCCGACACTGTTCCAGAAAGTGAGTGACAGCTCTACACCATCGG 52
|||||
69 SerAlaAlaHisCys.....LeuGlnGlnSerTyrIleValG1 81
53 GCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGGAGGAGGAGGAGGAGTGG 102
|||||
81 yLeuGlyLeuHisAsnLeuLysGlySerGlnGlnProGlySerArgMetL 98
103 TGGAGGCCACGCTCTCCGTCAGGCGACCCAGAGTACACAGACCCCTGCTC 152
|||||
98 euGlnAlaHisLeuSerIleGlnHisProAsnProAsnProSerPhe 114
153 GCTAACGACCTCATGCTCATCAAGTTGAGCAAGATCCGTGTCGAGTCTGA 202
|||||
115 AlaAsnAspLeuMetLeuIleLysLeuAsnGlnSerValIleGlnSerAs 131
203 CACCATCCGAGACATCAGATTGCTTCGACAGTGCCTACCGGGGAAGT 252
|||||
131 nThrIleArgSerIleProValAlaThrGlnCysProThrProGlyAspT 148
253 CTGCGCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGGTGAGCTCAGCGGT 302
|||||
148 hrcysLeuValSerGlyTyrPylGlnLeuLysAsnGly..Lys..... 161
303 GTGTGTGCTGCCCTCTTCAAGAGGTCCTCTGCCCATCGGGGGGCTGAC 352
161 .....
353 CCAGAGCTCTGCGTCCAGGAGGAATGCTTACCGTCTGAGTGGCTGAA 402
162 .....LeuProSerLeuLeuGlnCysValAs 170
403 CGTGTGCGTGTGTCTGANGAGGCTCTGCANTAAAGCTCTATGACCCGCTGT 452
|||||
```



```

170 nleuserValAlaSerGluGluIurhCysArgLeuLeuTyrAspProValT 187
453 ACCACCCCATCATGTTCTGCGCCGGCGAGGGCAAGACCAGAAGACTCC 502
||||| :::|||
187 yrh1leuserMetPheCysAlaGlyGlyGlnAspGlnLysAspser 203
503 TGCAACGTGAGAGAGGG 520
||||| :::|||
204 CysAsnGlyAspSerGly 209

seq_name: sp_human:P78525

seq_documentation_block:
ID P78525 PRELIMINARY; PRT; 666 AA.
AC P78525; P78526;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MYB PROTO-ONCOGENE PROTEIN (C-MYB).
GN C-MYB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER, AND PLACENTA;
RA WESTIN E.H., GORSE K.M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U22376; AAB49034.1; -.
DR EMBL; U22376; AAB49037.1; -.
DR PROSITE; PS00037; MYB_1; 3.
DR PROSITE; PS00334; MYB_2; 3.
DR PFWAM; PF00249; myb_DNA-binding; 3.
KW Proto-oncogene; Nuclear protein; DNA-binding; Repeat;
KW Transcription regulation; Activator.
FT DNA_BIND 34 MYB.
FT DNA_BIND 87 MYB.
FT DNA_BIND 87 138 MYB.
FT DNA_BIND 139 189 MYB.
SEQUENCE 666 AA; 75525 MW; 2E666087 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 310.00 | Length: 97 |
| Ratio: | 3.827 | Gaps: 1 |
| Percent Similarity: | 83.505 | Percent Identity: 71.134 |

alignment_block:

US-09-030-606-174/rev x p78525 ..

Align seg 1/1 to: p78525 from: 1 to: 666

1447 TTTTGTGAGACAGAGTCCTTACTCTGTGCCCCAGCTGGAGTATAGTGTG 1398
||| ::|||::: ::| |||||::||| |||
556 PheThrGlnThrSerProValAlaAspAlaProThrGlyValGlnTyrH1 572
TGATCTCAACTCACTTCAACCTCTGCTCCCATATTCAAGCAATCTCT 1348
|||:::||||| ||||| ||||| |||||
572 SASPheGlySerLeuGlnProLeuProProGlyPheLysArgPheSerC 589
GCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCTGCCACCATATCCA 1298
||||| ||||| ||||| ||||| |||
589 ySLeuSerLeuProArgSerTyrAspTyrArgHisProProCArgPro 605
GCTAACTTTTGTATTTTGTAGTACAGACAGGATTTTCAACATTTTGGCCAGG 1297
||||| ||| |||||:::||||| ||| |||||
606 AlaAsn.PheGluPheLeuValGlnThrGlyPheLeuHisValGlyGlnA 622
CTGGCTTGAACCTCCTTACCTCAAGTAGATCTGGCTGGCCTC.GCCTGCCAA 1199
||||| ||||| ||||| ||||| |||||
622 IagIlyLeuGlnLeuLeuThrSerGlyAspLeuProAlaSerAlaSerGln 638
AGTGTGGGATTACAGGCATGAGCCACCCTGCCAGCCT 1160
||||| ||||| ||||| ||||| |||||

```

639 SerAlaArgIleThrGlyValSerHisArgAlaArgPro 651
seq_name: sp_human:014608

seq_documentation_block:
ID 014608 PRELIMINARY; PRT; 1079 AA.
AC 014608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE UBIQUITOUS TPR MOTIF, Y ISOFORM.
GN UTY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 98022381.
RL LAHN B.T., PAGE D.C.;
RT "Functional coherence of the human Y chromosome.";
RL Science 278:675-680(1997).
DR EMBL; AF000996; AAC51843.1; -.
DR PFAM; PF00515; TPR; 5.
SQ SEQUENCE 1079 AA; 118073 MW; D7A4F281 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 280.00 | Length: 85 |
| Ratio: | 3.836 | Gaps: 1 |
| Percent Similarity: | 85.882 | Percent Identity: 72.941 |

alignment_block:

US-09-030-606-174/rev x 014608

Align seg 1/1 to: 014608 from: 1 to: 1079

```

1414 GCTGGAGTATAGTGGTGTGATCTCAACTCACTTCAACCTCTGCCCTCCCAT 1365
|||||::: |||||||:::||||||| |||||
996 AlagIyMetGIntPrCySasPlenSerSerLeuGlnProProProG1 1012
1364 ATTCAGCAATTCCTCTGCTCAGCCCTCCCAAGTAGCTGGATTACAGGC 1315
|||||:::||||| |||||||:::|||||:::|||||:
1012 yPhelysArgPheserGHisLeuSerLeuProAnsSerTrpAsnTyArgH 1029
1314 GCCTGCCACCATATCCAGCTAACTTTTGATTTTGTAGTACAGACAGGATT 1265
:::||||| ||:::||||||||| ||:::|||||||
1029 IsIeuProSerCySProThrAsnPhCySilePhe.ValGluThrGlyPh 1045
1264 TCACCATTTTGGCCAGGCTGTGCTTGAACCTTACCTCAAGTGATCTGC 1215
||||||| ||||||| ||||||| |||||||::: |||
1045 eHISHisValGlyGlnAlaCySLeuGlnLeuLeuThrSerGlyGlyLeuL 1062
1214 CTGCCTC.GCCTCCCAAGTGTGGGATTACAGGCATGAGCCACCTGCC 1166
||| ||||||| ||||||| |||||||:::||||| |||
1062 euaIaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisIaIa 1078
1165 CAG 1163
:::
1079 Arg 1079
seq_name: sp_human:Q99770
seq_documentation_block:
ID Q99770 PRELIMINARY; PRT; 139 AA.
AC Q99770;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 15.4 KD PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]

```


RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RX MEDLINE; 96207227.
RA ANDERSSON B., WENTLAND M.A., RICAFFENTE J.Y., LIU W., GIBBS R.A.;
RT "A "double adaptor" method for improved shotgun library
construction."
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA YU W., ANDERSSON B., WORLEY K.C., MUZNY D.M., DING Y., LIU W.,
RICAFFENTE J.Y., WENTLAND M.A., LENNON G., GIBBS R.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U79260; AAB50206.1; -.
KW Hypothetical protein.
SQ SEQUENCE 139 AA; 15443 MW; 47FA63F2 CRC32;

alignment_scores:
Quality: 270.50 Length: 94
Ratio: 3.559 Gaps: 2
Percent Similarity: 80.851 Percent Identity: 62.766

alignment_block:

US-09-030-606-174/rev x Q99770 ..

Align seg 1/1 to: Q99770 from: 1 to: 139

1459 TTTT TTTT TTTT TTTT TTTGAGACAGAGTCTTACTGTGCCCCAGCTGG 1410
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
48 PheLeuPhePhePhePheGluThrGlnSerHisSerValThr.ArgLeuG 64

1409 AGTATAGTGTGATCTCAACTCACTTCAACCTGTGCTGCCATATTCA 1360
|| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
64 LucysSerGlyThrIleSerAlaHisCysAsnLeuCysLeuProGlySer 80

1359 AGCAATTCTCTGCTGCTCAGCCTCCCAAGTAGCTGGATTACAGCGCCTG 1310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
81 SerAsnSerProAlaSerAlaSerArgValAlaGlyThrAlaGlyThrCys 97

1309 CCACCATATCCAGCTAACTTTGTATTTTAGTACAGACAGATTTCACC 1260
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
97 sArgArgAlaGlnLeuIlePheValPheLeuAlaGluMetGlyPheHisH 114

1259 ATTTGGCCAGGCTGCTTGAACCTCCTTACCTCAAGTGATCTGCTGCC 1210
|| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 lSvalGlyArgAspGlyLeuAspLeuAsnLeuValIleHisProProArg 130

1209 TCGCCTCCCAAGTGTGCTGGATTACAGGCA 1180
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 Ser...ProLysAlaLeuGlyLeuGlnAla 139

seq_name: sp_human:O60448

seq_documentation_block:

ID O60448 PRELIMINARY; PRT; 375 AA.
AC O60448;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE NEURONAL THREAD PROTEIN AD7C-NTP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-NEURONAL;
RA DE LA MONTE S.M., GHANBARI K., FREY W., BEHESHTI I., HAUSER S.A.,
GHANBARI H.A., WANDS J.R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF010144; AAC08737.1; -.
SQ SEQUENCE 375 AA; 41720 MW; 144FC2A2 CRC32;

alignment_scores:
Quality: 268.00 Length: 79
Ratio: 4.254 Gaps: 0
Percent Similarity: 79.747 Percent Identity: 64.557

alignment_block:

US-09-030-606-174/rev x O60448 ..

Align seg 1/1 to: O60448 from: 1 to: 375

1459 TTTT TTTT TTTT TTTT TTTGAGACAGAGTCTTACTGTGCCCCAGCTGG 1410
||| |||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
297 PheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaG 313

1409 AGTATAGTGTGATCTCAACTCACTTCAACCTGTGCTGCCATATTCA 1360
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 yValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeu 330

1359 AGCAATTCTCTGCTCAGCCTCCCAAGTAGCTGGATTACAGCGCCTG 1310
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
330 ysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeu 346

1309 CCACCATATCCAGCTAACTTTGTATTTTAGTACAGACAGATTTCACC 1260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 ProProHisProAlaAsnPheCysIlePheIleArgGlyGlyValSerPr 363

1259 ATTTGGCCAGGCTGCTTGAACCTCCTTACCTCAAG 1223
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
363 cTyrlLeuSerGlyTrpSerGlnThrProAspLeuArg 375

seq_name: sp_human:Q15662

seq_documentation_block:

ID Q15662 PRELIMINARY; PRT; 368 AA.
AC Q15662;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DE TRANSFORMATION-RELATED PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EPITHELIAL;
RA SHEN H., STEINBERG M.L.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24521; AAA36776.1; -.
FT NON_TER 1
SQ SEQUENCE 368 AA; 42029 MW; E4772637 CRC32;

alignment_scores:
Quality: 257.00 Length: 434
Ratio: 1.332 Gaps: 18
Percent Similarity: 44.470 Percent Identity: 26.728

alignment_block:

US-09-030-606-174/rev x Q15662 ..

Align seg 1/1 to: Q15662 from: 1 to: 368

1440 AGACAGAGTCTTACTGTGCCCCAGCTGAGATATAGTGATCTC 1391
||| ||||| ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
2 ArgLeuSerLeu...LeuSerProArgLeuGluCysAsnGlyMetIleLe 17

1390 AACTCACTTCAACCTTGCCCTCCATA.TTCAAGCAATTCTCTGCTCA 1342
:::||||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
17 uAlaHisCysLysLeuArgLeuProGlyPheLysArgPheSerCysLeuS 34

1341 GCCTCCCAAGTAGCTGGATTACAGGCGCTGCCACCATATCCAGCTAAC 1292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

KW Signal: Hydrolase.
FT SIGNAL 1 13 POTENTIAL.
FT CHAIN 14 242 PROTRYP SIN.
SQ SEQUENCE 242 AA; 26201 MW; DDD27DCA CRC32;

alignment_scores:
Quality: 221.00 Length: 216
Ratio: 1.826 Gaps: 10
Percent Similarity: 56.019 Percent Identity: 30.093

alignment_block:
US-09-030-606-174 x Q92099 ..

Align seg 1/1 to: Q92099 from: 1 to: 242

3 TCAGCCGCACACTGTTTCCAGAGTGAGTGACAGAGCTCTACACCATCGG 52
|||||
57 SerAlaAlaHisCysTyrLysSerArgValGlu.....ValAr 69
53 GGTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGG 102
:::| | | | |
69 gMetGlyGluHisHisIleArgVal...ThrGluGlyGlyGluGlnPheI 85
103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTGCTC 152
:::| | | | |
85 leSerSerSerArgValIleArgHisProAsnTyrSerTyrAsnIle 101
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCTGA 202
| | | | |
102 AspAsnAspIleMetLeuIleLysLeuSerLysProAlaThrLeuAsnGl 118
203 CACCATCCGGAGCATCAGCATTTGCTGCGAGTGCCCTACCGGGGGAAGT 252
: | | | | |
118 nTyrValGlnAlaValAlaLeuProSerSerCysAlaProAlaGlyThrM 135
253 CTGCGCTGTTTCTGCTGGGCTGCTGCTGCGAAGCGTGAGCTCACGGGT 302
| | | | |
135 etCysThrValSerGlyTyrPgl..... 142
303 GTGTGTGTCGCCCTCTTCAAGAGAGTCTCTGCCAGTCCGGGGGCTGAC 352
| | | | |
143SerThrGlnSer..... 146
353 CCAGAGCTGTGCGTCCCGAGGACAGATGCCCTACCGTCTGCGAGTGAA 402
| | | | |
147 ...SerSerAlaAsp..GlyAsnLys.....LeuGlnCysLeuAs 158
403 CGTGTGCGTGTCTGTGANGAGGCTGTGCANTAAAGCTCTATGACCCGCTGT 452
| | | | |
158 nIleProIleLeuSerAspArgAspCysAspAsnSerTyrProGlyMetI 175
453 ACCACCCCANCATGTCTGCGCGCGCGGGAGGGCAAGACCAAGAGACTCC 502
: | | | | |
175 leThrAspAlaMetPheCysAlaGlyTyrLeuGlnGlyGlyLysAspSer 191
503 TGCACGCTGAGAGAGGGGAAAGGGGAGGCGAGGCGACTCAGGGAAGG.. 550
| | | | |
192 Cys.....GlnGlyAspSerGlyGlyProValValCysAsnGlyG1 205
551TGAGAGAGGGGG.....AGACAGAGACACA 575
| | | | |
205 uLeuGlnGlyValValSerTyrPglTyrGlyCysAlaGlnArgAspHisP 222
576 CAGGCGCGCATGCCGAGATG.....CAGAGATGGAGAGACACA 613
| | | | |
222 roGlyValTyrAlaLysValCysLeuPheAsnAspTyrPleuGluThr 237

seq_name: sp_vertibrate:O93265

seq_documentation_block:
ID O93265 PRELIMINARY; PRT; 250 AA.
AC O93265;

DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE TRYP SIN GEN 1 PRECURSOR (EC 3.4.21.4).
GN TRP1.
OS Pleuronectes americanus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Pleuronectiformes; Pleuronectoidei; Pleuronectidae; Pleuronectes.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RA DOUGLAS S.E., GALLANT J.W.;
RT "Isolation of cDNAs for Trypsinogen from the Winter Flounder,
RT Pleuronectes americanus.";
RL J. Mar. Biotechnol. 0:0-0(1998).
DR EMBL; AF012462; AAC32751.1; ..
DR PFAM; PF00089; trypsin; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 23 250 TRYP SIN GEN 1.
SQ SEQUENCE 250 AA; 27466 MW; D387394D CRC32;

alignment_scores:
Quality: 220.00 Length: 192
Ratio: 2.018 Gaps: 7
Percent Similarity: 56.771 Percent Identity: 31.771

alignment_block:
US-09-030-606-174 x O93265 ..

Align seg 1/1 to: O93265 from: 1 to: 250

3 TCAGCCGCACACTGTTTCCAGAGTGAGTGACAGCTCTACACCATCGG 52
| | | | |
59 SerValAlaHisCysTyrTyrAsn.....ProTyrAlaMetGlnValMe 73
53 GGTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGG 102
| | | | |
73 tLeuGlyGluHisAsnLeuArgVal...PheGluGlyThrGluGlnLeuM 89
103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTGCTC 152
: | | | | |
89 etLysThrAspThrIleLeuTyrHisProAsnTyrAspTyrGlnThrLeu 105
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCTGA 202
| | | | |
106 AspPheAspIleMetLeuIleLysLeuTyrHisProValGlnValThrGl 122
203 CACCATCCGGAGCATCAGCATTTGCTGCGAGTGCCCTACCGGGGGAAGT 252
: | | | | |
122 uAlaValAlaProIleSerLeuProSerSerCysProValGlyGlyThrP 139
253 CTGCGCTGTTTCTGCTGGGCTGCTGCTGCGGAACGGTGAGCTCACGGGT 302
| | | | |
139 rocYserValSerGlyTyrPgl..... 146
303 GTGTGTGTCGCCCTCTTCAAGAGAGTCTCTGCCAGTCCGGGGGCTGAC 352
| | | | |
147AsnThrAlaArgAsp..GlyAspAsp 154
353 CCAGAGCTGTGCGTCCCGAGGACGAATGCCCTACCGTCTGCGAGTGCGTGA 402
| | | | |
155 Val.....TyrMetProThrLeuLeuGlnCysMetAs 165
403 CGTGTGCGTGTCTGTGANGAGGCTGTGCANTAAAGCTCTATGACCCGCTGT 452
: | | | | |
165 pValProIleIleAspGluGlnGlnCysMetLysSerTyrProGlyMetI 182
453 ACCACCCCANCATGTCTGCGCGGCGGAGGCAAGACCAAGAGAGACTCC 502
| | | | |
182 leSerProArgMetValCysAlaGlyPheMetAspGlySerArgAspAla 198

503 TGCAACGTGAGAGAGGG.....AAAGGGAGGGCAGGGC 537
||||| :||| :|||
199 CysasnGlyAspSerGlySerProLeuValCysArgGlyValThrGl 215
538 ACTCAGGAGAGGTGAGAGAGGGG 562
||| :||| :|||
215 yLeu..ValSerTrpGlyGlnGly 222

seq_name: sp_vertebrate:Q91515

seq_documentation_block:

ID Q91515 PRELIMINARY; PRT; 237 AA.
AC Q91515;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPSINOGEN (FRAGMENT).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG K., GAN L., LEE I., ROACH J., HOOD L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25747; AAA75001.1; -.
DR PFM; PF00089; trypsin; 1.
FT NON_TER 1 1
SQ SEQUENCE 237 AA; 25726 MW; 6A296989 CRC32;

alignment_scores:

Quality: 219.50 Length: 173
Ratio: 2.173 Gaps: 4
Percent Similarity: 58.382 Percent Identity: 30.636

alignment_block:

US-09-030-606-174 x Q91515 ..

Align seg 1/1 to: Q91515 from: 1 to: 237

3 TCAGCCGACACTGTTCACAGAGTGCAGAGCTCCTACACCATCGG 52
||||| :||| :|||
52 SerAlaAlaHisCysTyr.....LysSerArgValValAla 64
53 GCTGGGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGAGCCAGATGG 102
||||| :||| :|||
64 gLeuGlyGlnHisAsnIleArgAlaAsn...GluGlyThrGluGlnPheI 80
103 TGGAGGCCAGCTCTCCGTACGCGACACCCAGAGTACAACAGACCTTGCTC 152
:||| :||| :|||
80 leSerSerSerArgValIleArgHisProAsnTyrSerSerTyrAsnIle 96
153 GCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTGCCAGTCTGA 202
||||| :||| :|||
97 AspAsnAspIleMetLeuIleLeuLeuSerLysProAlaThrLeuAsnGl 113
203 CACCATCCGAGCATCAGCATGTGCTGCGAGTGCCTACCGCGGGGAACT 252
:||| :||| :|||
113 nTyrValGlnProValAlaLeuProSerSerCysAlaAlaIleGlyThrM 130
253 CTGCGCTCGTTCTGCGTGGGTCTGCTGGCGAAGCGGTGAGCTCAGGGGT 302
||| :||| :|||
130 etCysLysValSerGlyTrpGly..... 137
303 GTGTGTCTGCCCTCTCTCAAGAGGCTCTCTGCCAGTGGCGGGGCTGAC 352
137 137
353 CCAGAGCTCTGCGTCCAGGAGAAATGCTTACCGTGTGACAGTCCGTGAA 402
:||| :||| :|||
138 .AsnThrMetSerSerThrAlaAsp..ArgAsnLysLeuGlnCysLeuAs 153

403 CGTGTGCGTGTCTCTGANGAGGTCTGCANTAGCTCTATGACCCGCTGT 452
||| :||| :|||
153 nIleProIleLeuSerAspArgAspCysGlnAsnSerTyrProGlyMetI 170
453 ACCACCCANCATGTTCTGCGCGCGCGGAGGGCAAGACCAAGAGACTCC 502
:||| :||| :|||
170 leThrAspAlaMetPheCysAlaGlyTyrLeuGlnGlyLysAspSer 186
503 TGCAACGTGAGAGAGGGG 520
||| :|||
187 CysGlnGlyAspSerGly 192

seq_name: sp_vertebrate:Q92046

seq_documentation_block:

ID Q92046 PRELIMINARY; PRT; 249 AA.
AC Q92046; Q13261;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PREPROTRYPsin PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
OS Dissostichus mawsoni.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Perciformes; Notothenioidae; Nototheniidae; Dissostichus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97268652.
RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
RT "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in
RT Antarctic notothenioid fish.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3811-3816(1997).
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-I-XAA, LYS-I-XAA.
DR EMBL; U58945; AAB57732.1; -.
DR EMBL; U58835; AAB57728.1; -.
DR PFM; PF00089; trypsin; 1.
KW signal; Hydrolase.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 249 PROTRYPsin.
SQ SEQUENCE 249 AA; 27169 MW; 55C59EB4 CRC32;

alignment_scores:

Quality: 219.00 Length: 192
Ratio: 2.066 Gaps: 5
Percent Similarity: 55.208 Percent Identity: 32.292

alignment_block:

US-09-030-606-174 x Q92046 ..

Align seg 1/1 to: Q92046 from: 1 to: 249

3 TCAGCCGACACTGTTCACAGAGTGCAGAGCTCCTACACCATCGG 52
||||| :||| :|||
59 SerValAlaHisCysTrpTyrAsn.....ProTyrSerMetGlnValI 73
53 GCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
||||| :||| :|||
73 eLeuGlyAspHisAsnLeuArgVal...PheGluGlyThrGluGlnLeuM 89
103 TGGAGGCCAGCTCTCCGTACGCGACACCCAGAGTACAACAGACCTTGCTC 152
:||| :||| :|||
89 etLysThrAsnThrIleIleThrHisProSerTyrAspTyrGlnThrLeu 105
153 GCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTGCTCCAGTCTGA 202
||||| :||| :|||
106 AspPheAspIleMetLeuIleLysLeuTyrHisProValGluValThrGl 122


```
203 CACCATCCGGAGCATCAGCATTCCTTCGAGTGCCTTACCGGGGAACT 252
      :::::  |||  ::  :::::|||||  ::|||  |
122 uAlaValAlaProIleProIleuProThrSerCysProTyrGlyGlyLeuS 139
253 CTTCGCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGCGTGAGCTCAGCGGGT 302
      |||||  |||||  |||||  ::::  |||
139 erCysSerValSerGlyTyrGlyLeuAlaLysLeuGly..... 151
303 GTGTGTCTGCCCTCTTCAAGAGAGGTCTCTGCCCCAGTCGGCGGCTGAC 352
151 ..... 151
353 CCAGAGCTCTGCGTCCCGAGGAGA.ATGCCTACCGTGCTGAGTGCCTGA 401
      :::::||||  |||||  :::::|||||  ::|||  |
152 .....GlyGluAlaTyrMetProThrLeuLeuGlnCysLeuA 164
402 ACGTGTCCGGTGTGTCTGANGAGGTCTGCANTAGCTCTATGACCCGCTG 451
      |||||  ::|||  ::|||  ::|||  ::|||  |||
164 snValProIleValAspGlnGlnValCysGluAsnThrTyrProGlyLeu 180
452 TACCACCCCANCATGTCTGCGCCCGCGGAGGGCGAAGCAGACGAGACTC 501
      ::|||  |||||  ::  |||||  ::|||  |||||  ::
181 lIeSerThrThrMetValCysAlaGlyTyrMetGlnGlyGlyLysAspAl 197
502 CTGCACGCTGAGAGAGGGGAAA.....GGGAGGGCAGGC 536
      :|||||  ::|||  ::|||  |||||  ::|||  ::
197 aCysAsnGlyAspSerGlySerProLeuValCysAspGlyGluValGlnG 214
537 GACTCAGGGAAGGTTGAGAGGGGG 562
      |||  ::|||  ::|||
214 lYLeu...ValSerTyrGlyGlnGly 221
seq_name: sp_vertebrate:093266
```

```
seq_documentation_block:
ID 093266 PRELIMINARY; PRT; 242 AA.
AC 093266;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
DE TRYPSINOGEN 2 PRECURSOR (EC 3.4.21.4).
GN TRP2.
OS Pleuronectes americanus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Pleuronectiformes; Pleuronectoidae; Pleuronectidae; Pleuronectes.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE, PYLORIC CAECA, AND PANCREAS;
RA DOUGLAS S.E., GALLANT J.W.;
RT "Isolation of cDNAs for Trypsinogen from the Winter Flounder,
RT Pleuronectes americanus."
RL J. Mar. Biotechnol. 0:0-0(1998).
DR EMBL; AF012463; AAC32752.1; -.
DR PFAM; PF00089; trypsin; 1.
KW SIGNAL; Hydrolase.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 21 242 TRYPSINOGEN 2.
SQ SEQUENCE 242 AA; 26180 MW; 7E72C91C CRC32;
```

```
alignment_scores:
  Quality: 213.50      Length: 202
  Ratio: 1.941        Gaps: 7
Percent Similarity: 54.455 Percent Identity: 28.713
```

alignment_block:

us-09-030-606-174 x 093266 ..

Align seg 1/1 to: 093266 from: 1 to: 242

3 TCAGCGGCACACTGTTCCAGAGTGAAGTGACAGAGCTCTACACCATCGG 52

```
|||||  |||||  :::::  |||::  ::
57 SerAlaAlaHisCysTyrLysSerArgValGlu.....ValAr 69
53 GTGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGG 102
      ::|||  |||  :::::  :::::  |||  ::|||  ::|||  |
69 gmetGlyGlnHisLysIleArgValAsn...GlnGlyThrGlnPheV 85
103 TGGAGGCCAGCCCTCCGTACGGCACCCAGAGTACACAGACCCCTGTGCTC 152
      ||:::  |||  ::|||  |||||  ::|||  ::
85 alSerSerArgValIleArgHisProAsnTyrAspSerTyrAsnIle 101
153 GCTAACGACCTCATGTCTCATCAAGTTGACGAATCCGTGCCAGTCTGA 202
      |||||  ::|||  |||||  |||||  :::::  :::::  :::::
102 AspAsnAspIleMetLeuIleLysLeuSerLysProAlaThrLeuAsnGl 118
203 CACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCTTACCGCGGGAAC 252
      :  :::::  :::::  ::|||  |||||  |||||  ::
118 nTyrValLysThrValAlaLeuProSerSerCysAlaProAlaGlyThrM 135
253 CTTCGCTCGTTTCTGGCTGGGCTCTGCTGGCGAAGCGTGAGCTCAGGGGT 302
      |||  |||||  |||||  |||||  |||||  |||
135 etCysLysValSerGlyTyrGly..... 142
303 GTGTGTCTGCCCTCTTCAAGAGAGGTCTCTGCCAGTCGCGGGGCTGAC 352
      ::  |||||  :::::  |||
143 .....AsnThrMetSerSerAlaAspAsn..Gly..... 151
353 CCAGAGCTCTGCGTCCCGACGACAGAATGCCCTACCGTGTGACAGTGC 402
152 ..... 158
403 CGTGTCCGGTGTGTCTGANGAGGTCTGCANTAGCTCTATGACCCGCTGT 452
      ::::  ::|||  ::  |||  ::|||  ::|||  ::
158 pIleProIleLeuSerPheSerAspCysAsnAsnAlaTyrProGlyMetI 175
453 ACCACCCCANCATGTCTGCGCCCGCGGAGGGCGAAGCAGACGAGACTCC 502
      ::|||  |||||  |||||  ::  |||||  |||||  |||
175 lThrAspSerMetPheCysAlaGlyTyrLeuGlnGlyGlyLysAspSer 191
503 TGCAACGTGAGAGAGGGAAGGGGAGGCGGAGCTCAGGGAAGGGTG 552
      ||:::  ::|||
192 CysGlnGlyAspSerGly..... 197
553 GAGAAGGGGAGAGACAGACACAGGCGG.....CATGGCGAGA 593
198 .....GlyProValIleCysAsnGlyGluL 206
594 TGCAG 598
      ::|||
206 eugIn 207
```

seq_name: sp_vertebrate:093594

```
seq_documentation_block:
ID 093594 PRELIMINARY; PRT; 178 AA.
AC 093594;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
DE TRYPSIN (EC 3.4.21.4) (FRAGMENT).
OS Dicentrarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Perciformes; Percoidae; Moronidae; Dicentrarchus.
RN [1]
RP SEQUENCE FROM N.A.
RA PERES A., ZAMBONINO INFANTE J.L., CAHU C.L.;
RT "Dietary regulation of activities and mRNA levels of trypsin and
RT amylase in sea bass (Dicentrarchus labrax) larvae."
RL Fish Physiol. Biochem. 19:145-152(1998).
DR EMBL; AF006882; CAA07315.1; -.
DR PFAM; PF00089; trypsin; 1.
KW Hydrolase.
```


FT NON_TER 1 1
FT NON_TER 178 178
SQ SEQUENCE 178 AA; 19352 MW; 506224A9 CRC32;

alignment_scores:

Quality: 212.50 Length: 202
Ratio: 1.986 Gaps: 6
Percent Similarity: 52.970 Percent Identity: 27.723

alignment_block:

us-09-030-606-174 x 093594 ..

Align seg 1/1 to: 093594 from: 1 to: 178

```
3 TCAGCCGACACTGTTCAGAGTAGTGACAGAGCTCCTACACATCGG 52
|||||
23 SerAlaAlaHisCysTyrLysSerArgValGlu.....ValAr 35
53 GCTGGCCCTGCACAGTCTTGAGCCCGACCAAGAGCCAGGAGCCAGATGG 102
|||||
35 glenGlyGluHisAsnIleArgVal..ThrGluAsnThrGluGlnPheI 51
103 TGGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCTC 152
:
51 leSerSerSerArgValIleArgHisProArgTyrSerSerTyrAsnIle 67
153 GCTAACGACCTCATGTCTCATCATGTGACAGCAATCCGTGCCAGTCTGA 202
|||||
68 AspAsnAspIleMetIleuIleuLysLeuSerLysProAlaThrLeuAsnG1 84
203 CACCATCCGGAGACATCAGCATGCTTCGACAGTGCCTACCGCGGGAAGT 252
:
84 nTyrValGlnProValAlaLeuProThrSerCysAlaProAlaGlyThrM 101
253 CTGGCCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGGTGAGCTCAGGGGT 302
|||
101 etCysThrValSerGlyTyrGly..... 108
303 GTGTGTCTGCCCTCTTCAAGAGAGGTCTCTGCCCAGTCGGGGGGCTGAC 352
108 ..... 108
353 CCAGAGCTTCGCGTCCCGACGAGAATGCCTACCGCTGCAGTGCAGTGA 402
:
109 .AsnThrMetSerSerThrAlaAsp..ArgAsnLysLeuGlnCysLeuAs 124
403 CGTGTGCGGTGTCTGTGANGAGTCTGCANTAAAGCTATAGCCGCTGT 452
|
124 nIleProIleLeuSerPheLysAspCysAspAsnSerTyrProGlyMetI 141
453 ACCACCCCANCATGTCTGCGCGCGCGGAGGGCAAGACCAAGAAAGACTCC 502
:
141 leThrAspAlaMetPheCysAlaGlyTyrLeuGlnGlyGlyLysAspSer 157
503 TGCACGCTGAGAGAGGGGAAAGGGGAGGGCAGGCGACTCAGGGAGGGGTG 552
|
158 CysGlnGlyAspSerGly..... 163
553 GAGAGGGGGAGAGACAGAGACACAGGCGCG.....CATGGCGAGA 593
164 .....GlyProValValCysAsnGlyGluL 172
594 TGCAG 598
172 eugln 173
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:36 ; Search time 1809.22 seconds
(without alignments)
1590.700 Million cell updates/sec

Title: US-09-030-606-174

Perfect score: 1459

Sequence: 1 GGTCAGCCGCACACTGTTC.....TCAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| 1 | 358.4 | 24.6 | 360 | 50 | AI675523 | AI675523 wc01f01.x |
| 2 | 286.4 | 19.6 | 777 | 48 | AI557281 | AI557281 PT2.1.15 |
| 3 | 245.4 | 16.8 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 4 | 244.6 | 16.8 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 5 | 242.8 | 16.6 | 722 | 48 | AI557025 | AI557025 PT2.1.10 |
| 6 | 226.2 | 15.5 | 481 | 50 | AI669421 | AI669421 ty32d03.x |
| 7 | 225.4 | 15.4 | 415 | 48 | AI609972 | AI609972 tt78c10.x |
| 8 | 223 | 15.3 | 407 | 35 | AA535216 | AA535216 nj75f05.s |
| 9 | 222.4 | 15.2 | 479 | 26 | W96522 | W96522 ze43f08.r1 |
| 10 | 222.2 | 15.2 | 437 | 36 | AA644090 | AA644090 ab62b04.s |
| 11 | 220.2 | 15.1 | 500 | 36 | AA626040 | AA626040 af94a08.s |
| 12 | 220.2 | 15.1 | 454 | 49 | AI634187 | AI634187 ts55a08.x |
| 13 | 220 | 15.1 | 374 | 38 | AA746911 | AA746911 nx69c10.s |
| 14 | 219.8 | 15.1 | 434 | 36 | AA630854 | AA630854 nt57f09.s |
| 15 | 219.6 | 15.1 | 388 | 39 | AA837686 | AA837686 oe05h09.s |
| 16 | 219.4 | 15.0 | 323 | 35 | AA536040 | AA536040 nj81c09.s |
| 17 | 219.2 | 15.0 | 484 | 33 | AA410788 | AA410788 zt35b11.r |
| 18 | 219.2 | 15.0 | 380 | 36 | AA613624 | AA613624 no18d03.s |
| 19 | 218.4 | 15.0 | 458 | 51 | AI733856 | AI733856 zo19c03.y |
| 20 | 218.2 | 15.0 | 408 | 34 | AA486877 | AA486877 ab16h04.s |
| 21 | 218 | 14.9 | 533 | 39 | AA833875 | AA833875 od64e08.s |
| 22 | 217.2 | 14.9 | 470 | 34 | AA456924 | AA456924 aa90b09.s |
| 23 | 217.2 | 14.9 | 329 | 36 | AA643770 | AA643770 np06e10.s |
| 24 | 217.2 | 14.9 | 466 | 42 | AI087040 | AI087040 oy70a10.x |
| 25 | 217.2 | 14.9 | 365 | 46 | AI419337 | AI419337 tf27h01.x |
| 26 | 217.2 | 14.9 | 397 | 46 | AI421950 | AI421950 tf45d05.x |
| 27 | 217 | 14.9 | 415 | 34 | AA515048 | AA515048 ng67h10.s |
| 28 | 216.6 | 14.8 | 461 | 30 | AA225406 | AA225406 nc24d02.r |
| 29 | 216.6 | 14.8 | 301 | 39 | AA828592 | AA828592 od74e10.s |
| 30 | 216.4 | 14.8 | 528 | 39 | AA833896 | AA833896 od64g08.s |
| 31 | 216.4 | 14.8 | 344 | 41 | AI054030 | AI054030 q166a12.x |
| 32 | 216.2 | 14.8 | 428 | 41 | AI066646 | AI066646 oz82d10.x |
| 33 | 216 | 14.8 | 569 | 53 | HSM007223 | AI042373 Homo sapi |
| 34 | 215.6 | 14.8 | 367 | 45 | AI361090 | AI361090 qy04d06.x |
| 35 | 215.6 | 14.8 | 524 | 47 | AI524240 | AI524240 fh11d12.x |
| 36 | 215.4 | 14.8 | 457 | 31 | AA284247 | AA284247 zc65d01.T |
| 37 | 215.4 | 14.8 | 480 | 49 | AI635028 | AI635028 tz03d06.x |
| 38 | 215.2 | 14.7 | 376 | 29 | AA176978 | AA176978 nc01c01.s |
| 39 | 215.2 | 14.7 | 474 | 46 | AI457313 | AI457313 tl173a05.x |
| 40 | 215 | 14.7 | 440 | 36 | AA601278 | AA601278 no15b09.s |
| 41 | 214.8 | 14.7 | 624 | 36 | AA601356 | AA601356 no16b05.s |
| 42 | 214.8 | 14.7 | 441 | 29 | HSM003412 | AI038936 Homo sapi |
| 43 | 214.6 | 14.7 | 441 | 29 | AA169245 | AA169245 zp19d03.s |
| 44 | 214 | 14.7 | 357 | 45 | AI340832 | AI340832 tb55a10.x |
| 45 | 214 | 14.7 | 413 | 46 | AI433104 | AI433104 th41d12.x |

ALIGNMENTS

RESULT 1
AI675523/c
LOCUS AI675523 360 bp mRNA
DEFINITION wc01f01.x1 NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2313913 3',
ACCESSION AI675523
NID g4876003
VERSION AI675523.1 GI:4876003

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 360)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT On May 9, 1996 this sequence version replaced gi:1133152.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 357.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .360 |

```

/clone="IMAGE:2313913"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."

```

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 24.6% | Score 358.4; | DB 50; | Length 360; |
| Best Local Similarity | 99.7%; | Pred. No. 3.4e-57; | | |
| Matches 359; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 636 | GAGAGAACTGAGAGAAACAGAGAAATTAACACAGGAATTAAGAGAAGCAAGGAAGAGAGA | 693 |
| | | | |
| Db | 360 | GAGAGAACTGAGAGAAACAGAGAAATTAACACAGGAATTAAGAGAAGCAAGGAAGAGAGA | 301 |
| QY | 696 | GAAACAGAAACAGACATGGGGAGGCAGAAACACACACACATAGAAATGCAGTTGACCTTC | 755 |
| | | | |
| Db | 300 | GAAACAGAAACAGACATGGGGAGGCAGAAACACACACACATAGAAATGCAGTTGACCTTC | 241 |
| QY | 756 | CAACAGCATGGGCGCTGAGGGCGGTGACCTCCACCCAATAGAAATCCTCTATACTTT | 815 |
| | | | |
| Db | 240 | CAACAGCATGGGCGCTGAGGGCGGTGACCTCCACCCAATAGAAATCCTCTATACTTT | 181 |
| QY | 816 | TGACTCCCCAAAACCTGACTAGAAATAGCCTACTGTGACGGGGAGCCTTACCAATAAC | 875 |
| | | | |
| Db | 180 | TGACTCCCCAAAACCTGACTAGAAATAGCCTACTGTGACGGGGAGCCTTACCAATAAC | 121 |
| QY | 876 | ATAAATAGTCGATTTATGCATACGTTTATGCATTTCATGATATACCCTTTGTTGGAATTTT | 935 |
| | | | |
| Db | 120 | ATAAATAGTCGATTTATGCATACGTTTATGCATTTCATGATATACCCTTTGTTGGAATTTT | 61 |

| QY | 936 | TTGATATTTCTAAGCTACACAGTTCGTCGTGAATTTTAAATGTTGCAACTCTCT | 995 |
|----|-----|---|-----|
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 60 | TTGATATTTCTAAGCTACACAGTTCGTCGTGAATTTTAAATGTTGCAAAATCTCT | 1 |

| | |
|------------|--|
| RESULT | 2 |
| A1557281 | |
| LOCUS | A1557281 777 bp mRNA |
| DEFINITION | PT2.1-15_G12.r tumor2 Homo sapiens CDNA 3', mRNA sequence. |
| ACCESSION | A1557281 |
| NID | g4489644 |
| VERSION | A1557281.1 GI:4489644 |

| | |
|----------|--------|
| KEYWORDS | EST. |
| SOURCE | human. |

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 777)

AUTHORS Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J. and Hood, L.

TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)

COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Mathew Huang

Leroy

University of Washington

Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195

Washington, Sea
 106380100

Tel: 5106280100
Fax: 5106280708

Fax: 3106280106
Email: huanqian@yahoo.com

Email: ivan@ngilleyalbo.com
Location/Qualifit

| | |
|----------|---------------------|
| FEATURES | Location/Qualifiers |
| source | 1. .777 |

Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."

```

| BASE COUNT ORIGIN | 162 a | 205 c | 227 g | 157 t | 26 others |
|----------------------|-------|-------|-------|-------|-----------|
|----------------------|-------|-------|-------|-------|-----------|

Query Match 19.6%; Score 286.4; DB 48; Length 777;

Best Local Similarity 79.88; Pred. No. 6.1e-44;
Matches 387; Conservative 0; Mismatches 14; Indels 84; Gaps 1.

| | | | |
|----|-----|--|-----|
| QY | 25 | AGTAGTGACAGAGCTCTTACACCATCGGGCTGGGCTTGACACAGTCTTGAGGGCCGACCAAG | 84 |
| Db | 111 | ACTGTTCCAGAACTCTTACACCATCGGGCTGGGCTTGACACAGTCTTGAGGGCCGACCAAG | 170 |
| QY | 85 | AGCCAGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGCACCCAGAGTACACAGAC | 144 |
| Db | 171 | AGCCAGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGCACCCAGAGTACACAGAC | 230 |
| QY | 145 | CCTTGCTCGTAACGACCTCATGCTCATCAAGTTGGACGAATCGTGTCGAGTCTGACA | 204 |
| Db | 231 | CCTTGCTCGTAACGACCTCATGCTCATCAAGTTGGACGAATCGTGTCGAGTCTGACA | 290 |
| QY | 205 | CCATCCGAGCATCAGCATTTGCTTCGCAGTGCCTTACCGCGGGGAACTCTTGCCCTGTTT | 264 |
| Db | 291 | CCATCCGAGCATCAGCATTTGCTTCGCAGTGCCTTACCGCGGGGAACTCTTGCCCTGTTT | 350 |
| QY | 265 | CTGGCTGGGGTCTGCTGGCGAAGCGGTGAGCTCAAGGGGTGTGTCTGCCCTCTTCAAGGA | 324 |
| Db | 351 | CTGGCTGGGGTCTGCTGGCGAAGCGGGTGTGTCTGCCCTCTTCAAGGA | 376 |
| QY | 325 | GGTCCCTTGCCCACTCGCGGGGGCTGACCCAGAGCTCTGCGTCCAGGACGAGATGCTTAC | 384 |
| Db | 376 | GGTCCCTTGCCCACTCGCGGGGGCTGACCCAGAGCTCTGCGTCCAGGACGAGATGCTTAC | 386 |

QY 385 CGTGTGACGTGCGTGAACGTGTGCTGTCGANGAGGCTTCGANTTAAGCTTATGA 444
|||||
Db 387 CGTGTGACGTGCGTGAACGTGTGCTGTCGANGAGGCTTCGANTTAAGCTTATGA 446
QY 445 CCCGCTGTACCCACCANCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGTCTCTG 504
|||||
Db 447 CCCGCTGTACCCACCACATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGTCTCTG 506
QY 505 CAACG 509
|||||
Db 507 CAACG 511

RESULT 3
AI686689/c 569 bp mRNA EST 27-MAY-1999
LOCUS tu35911.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253092 3'
DEFINITION similar to TR:Q92046 Q92046 PREPROTRYP SIN PRECURSOR ;, mRNA
sequence.

ACCESSION AI686689
NID 94897983
VERSION AI686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 569)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1. 569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

FEATURES
source
BASE COUNT 125 a 150 c 170 g 124 t
ORIGIN
Query Match 16.8%; Score 245.4; DB 50; Length 569;
Best Local Similarity 79.6%; Pred. No. 2.3e-36;

Matches 339; Conservative 0; Mismatches 4; Indels 83; Gaps 1;
QY 84 GAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGA 143
|||||
Db 569 GAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGA 510
QY 144 CCCTTGGCTCCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAGTCTGAC 203
|||||
Db 509 CCCTTGGCTCCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAGTCTGAC 450
QY 204 ACCATCCGAGCATCAGCATTTGCTTCGACAGTGCCTACCGCGGGGAACCTTGCCTGTT 263
|||||
Db 449 ACCATCCGAGCATCAGCATTTGCTTCGACAGTGCCTACCGCGGGGAACCTTGCCTGTT 390
QY 264 TCTGCTGGGCTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTCTTCAAGG 323
|||||
Db 389 TCTGCTGGGCTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTCTTCAAGG 365
QY 324 AGTCTCTGCCCCAGTCCGCGGGGCTGACCCAGAGCTCTCCGTCACAGCAGATGCCCTA 383
|||||
Db 365 -----GGCAGAAATGCCCTA 353
QY 384 CCGTGTGACGTGCGTGAACGTTGCGGTGTGTCTGANGAGGCTTCGANTTAAGCTTATG 443
|||||
Db 352 CCGTGTGACGTGCGTGAACGTTGCGGTGTGTCTGANGAGGCTTCGANTTAAGCTTATG 293
QY 444 ACCCGCTGTACCAACCCANCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGTCTCT 503
|||||
Db 292 ACCCGCTGTACCAACCCANCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGTCTCT 233
QY 504 GCAACG 509
|||||
Db 232 GCAACG 227

RESULT 4
AA551449 415 bp mRNA EST 05-SEP-1997
LOCUS nj55e05.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996416
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; , mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 415)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 640 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.
Location/Qualifiers
1. 415
FEATURES
source

ACCESSION AA535216
NID g2279469
VERSION AA535216.1 GI:2279469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 407)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 5, 1995 this sequence version replaced gi:798604.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquil,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbip/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
source Location/Qualifiers

1..407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone="IMAGE:998337"
/clone_lib="NCI-CGAP_Pr10"
/sex="male"
/tissue_type="invasive prostate tumor"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMP10; mRNA made from
invasive prostate tumor, CDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

BASE COUNT 112 a 93 c 123 g 79 t
ORIGIN

Query Match 15.3%; Score 223; DB 35; Length 407;
Best Local Similarity 86.3%; Pred. NO. 3.2e-32;
Matches 258; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

OY 1161 GGCTGGCAGGGTGGCTCATGCTGTAATCCACGACTTTGGAGGC-GAGGCAGGCAGA 1219
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 107 GGCTGGGCATGCTGCTCATGCTGTAATCCACGACTTTGAGAGGCTAAGGCAGGCAGA 166
OY 1220 TCACCTGAGGTAAGGATTCAGACCAAGCCTGGCCAAATGTTGAATCCTGTCTACT 1279
||| ||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 167 TCGCCTGAGTCCAGGGGTTCAGACCAAGCCTGGCCCAACATGTTGAACCCCGTCTACT 226
OY 1280 AAAAATACAAAAGTTAGCTGATATGTTGGCAGGCGCCTGTAATCCACGACTTGGGAG 1339
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 227 AAAAATACAAAATTAGCTGGGCACGGTGGCACGTCCTGTAATCCACGACTTGGGAG 286
OY 1340 GCTGAGCAGGAGAAATGCTGAATATGGAGGCAAGGTTGAAGTGAGTTGAGATCACA 1399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 287 GCTGAGCAGGAGAAATGCTGAACCCAGGAGCGGGAAGTTGCAGTGAGCCCAAGATCGTG 346
OY 1400 CCACATACTCCAGCTGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAA 1458
||||| ||||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 347 CCACATGCACTCCAGCTTGGGTGACAGAGTGAGACTCCGTCCTCAAAAAAAAAAAAAAAA 405

RESULT 9

W96522/c
LOCUS 479 bp mRNA EST 29-NOV-1996
DEFINITION ze43f08.r1 Soares retina N2b4HR Homo sapiens CDNA clone
IMAGE:361767 5' similar to contains Alu repetitive element; mRNA
sequence.

ACCESSION W96522
NID g1426448
VERSION W96522.1 GI:1426448
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 479)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636214.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3253 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 457.

FEATURES
source Location/Qualifiers

1..479
/organism="Homo sapiens"
/db_xref="GDB:1278470"
/db_xref="taxon:9606"
/map="1"
/clone="IMAGE:361767"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand CDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded CDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 89 a 134 c 100 g 152 t 4 others
ORIGIN

Query Match 15.2%; Score 222.4; DB 26; Length 479;
Best Local Similarity 83.3%; Pred. No. 4.1e-32;
Matches 264; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

OY 1144 TCAAGACTCTACAAGAAGGCTGGCAGGGTGCATGCTGTAATCCACGACTTTGGG 1203
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 338 TCAAAAAGATAAAGGAGAGCCGGGAGCAGAGTGCATGCTGTAATCCACGACTTTGGG 279

OY 1204 AGGC-GAGGAGGCAGATCACTTGAGGTTAAGAGTTCAAGACCAAGCCTGGCCAAATGCT 1262
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

[illegible]

| RESULT | 10 |
|-----------------|--|
| AA644090/c | |
| LOCUS | |
| DEFINITION | AA644090 437 bp mRNA EST 06-MAR-1998 |
| IMAGE:845359 3' | ab62b04.s1 Stratigene lung carcinoma 937218 Homo sapiens CDNA clone |
| sequence. | similar to contains Alu repetitive element; mRNA |
| ACCESSION | AA644090 |
| NID | 92569308 |
| VERSION | AA644090.1 |
| KEYWORDS | GI:2569308 |
| SOURCE | EST. |
| ORGANISM | human. |
| REFERENCE | Homo sapiens |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. |
| TITLE | 1 (bases 1 to 437) |
| JOURNAL | Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R. |
| COMMENT | WashU-NCI human EST Project Unpublished (1997) On Sep 19, 1997 this sequence version replaced gi:1520385. |

FEATURES

source

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 834 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 416.

LOCALIZATION/QUALITIES
1. .437

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
/clone="IMAGE:845359"
```

```
/clone_lib="Stratagene lung carcinoma 937218"
```

```
/tissue_type="lung carcinoma"
```

```
/cell_line="NCI-H69"
```

```
/dev_stage="cell line NCI-H69"
```

```

/lab_host="SOLR (kanamycin resistant)"

```

```
ECORT: site2: xbot: cloned unidirectionally prime
/notes: Organ: Lung; Vector: pBluescript SK-; site_1:
```

01190 dt: small cell carcinoma cell line NCI-H69, Average

Insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5'

[illegible]

98 a 112 c 78 g 149 t

BASE COUNT
ORIGIN

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 15.28; | Score 222.2; | DB 36; | Length 437; |
| Best Local Similarity | 85.58; | Pred. No. 4.5e-32; | | |
| Matches 259; Conservative | 0; | Mismatches 43; | Indels 1; | Gaps 1; |

| | | | |
|----|------|---|------|
| QY | 1158 | AGAGGCTGGGGCAGGGGTGGCTCATGCCGTGAATCCCAAGCACTTTGGGAGG-CGAGGCAGGC | 1216 |
| | | | |
| Db | 310 | ACAGGCTGGGTACAGGGGCTCATGCTTGTAAATCCCAAGCACTTTGGGAGGCCGAGGCGGGC | 251 |
| QY | 1217 | AGATCACTTGAGGTAAGGAGTTCAAGACCAGCCTGGCCAAAATGGTGAATCCTGTCTGT | 1276 |
| | | | |
| Db | 250 | AGATCACTTGAGATCAGAGAGTTGAGACCAGCCTGGCCAAACATGTTGAAACCCCTGTGTCT | 191 |
| QY | 1277 | ACTAAATAATACAAAAGTTAGCTGATATGTTGGCAGCGCCCTGTAATCCCACTACTTGG | 1336 |
| | | | |
| Db | 190 | ACTAAATAATACAAAATATTAGCCAGGTGTGTTGGCAGGTGCTGTAAATCCCACTACTCGG | 131 |
| QY | 1337 | GAGGCTGAGGCAGGAGAATGCTTGAATATGGAGGCAGAGGTTGAAGTGAGTTGAGATC | 1396 |
| | | | |
| Db | 130 | GAGGCTGAGGCAGGAGAATGCTTGAACCCCGGAGGTAAGGTTGCAGTGAGCCACAGATC | 71 |
| QY | 1397 | ACACCATACTACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAA | 1456 |
| | | | |
| Db | 70 | GCACCACTGCACCTCCAGCCCTGGGTGATAGAGTAAGACTCAGTCTCAAAAAAAAAAAAAA | 11 |
| QY | 1457 | AAA 1459 | |
| | | | |
| Db | 10 | AGA 8 | |

RESULT 11
AA626040/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA626040 500 bp mRNA EST 15-OCT-1997
af94a08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens CDNA clone
IMAGE:1055414 3' similar to contains Alu repetitive
element;contains element MER22 repetitive element ;, mRNA sequence
AA626040
g2538427
AA626040.1 GI:2538427
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 500)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
On Apr 14, 1993 this sequence version replaced gi:503350.

Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.11n1.gov) for
 more information. 40m13 End EM from Another

seq primer: -40m13 fwd. E11

FEATURES

Source

10

/organism="Homo sapiens"

```

/db_xref="taxon:9606"

```

```

/clone="IMAGE:1055414"
/cleanlib="Sources total" futuc nb3ue8 out

```

```
/dev stage="8-9 weeks"
```

```
/dev=0 0 0 0
/lab host="DH10B"
```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

(total) fetus material with a Not I - oligo(dT) primer [5'

TTGTTACCAATCTGAGTGGGAGCGGGCCCTTAATTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Donaldo. "

| BASE COUNT | 115 a | 123 c | 107 g | 154 t | 1 others |
|------------|-------|-------|-------|-------|----------|
| ORIGIN | | | | | |

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match | 15.1% | Score 220.2; | DB 36; | Length 500; |
| Best Local Similarity | 81.2%; | Pred. No. 1e-31; | | |
| Matches 255; Conservative | 0; | Mismatches 59; | Indels 0; | Gaps 0; |

[illegible]

RESULT 12
AI634187/c
LOCUS
DEFINITION
AI634187 454 bp mRNA EST 26-APR-1999
ts55a08.x1 NCI_CGAP_k1d8 Homo sapiens CDNA clone IMAGE:2232470 3
similar to contains Alu repetitive element; contains element TAR1
repetitive element ;, mRNA sequence.
AI634187
AI685517
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 454)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 18, 1998 this sequence version replaced gi:3138239.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

| FEATURES | High quality sequence stop: 411 |
|----------|---------------------------------|
| source | Location/Qualifiers |
| | 1. .454 |

```

source
1. .454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5"
/clone="IMAGE:2232470"
/clone_lib="NCI_CGAP_Kid8"
/tissue_type="renal cell tumor"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.2 kb. Life Technologies catalog #:
11524-014"

```

| | | | | |
|------------|-------|-------|------|-------|
| BASE COUNT | 118 a | 103 c | 93 g | 140 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match | 15.1%; | Score 220.2; | DB 49; | Length 454; |
| Best Local Similarity | 81.8%; | Pred. No. 1e-31; | | |
| Matches 266; Conservative | 0; | Mismatches 58; | Indels 1; | Gaps 1; |

| | | | |
|----|------|---|------|
| QY | 1130 | AGAACACAGGAAAAATCATCAGACTCTCTACAAAGAGGCTGGGCCAGGGCTGCCTCATGCTGTAAAT | 1189 |
| | | | |
| Db | 325 | AGTAAATTATATATATAAAATACCAAAAAAGAGGCTGGGCCGTGGCTCACGCCCTGTAAAT | 266 |
| QY | 1190 | CCCAGCACCCTTTGGGAGG-CGAGGCAGGCGAGATCACTTGAGGTAAGGAGTTCAGACCAGC | 1248 |
| | | | |
| Db | 265 | CCCAGCACCCTTTGGGAGGCGGAGGCAGCGATGATCACTAAGGTCAGGAGTTCGAGACCAGC | 206 |
| QY | 1249 | CTGGCCAAATAGGTGAATCCTGCTCTGTACTAAATAATCAAAAAGTTAGCTGGATATGGTG | 1308 |
| | | | |
| Db | 205 | CTGGCCAAATAGGTGAATCCTGCTCTGTACTAAATAATCAAAAAGTTAGCCGGGTGGTG | 146 |
| QY | 1309 | GCAGGCGCCTGTATCCAGCTACTTGGAGGCGTGAGCCAGGAGATTCGTTGATATGG | 1368 |
| | | | |
| Db | 145 | GCGTGCGCCTGTAGTCCAGCTACTCGGGAGGCGTGAGCCAGAGAATCGCTTGAACCCGG | 86 |
| QY | 1369 | GAGGCAGAGGTTGAAGTGAAGTTGAGATCACACCCTATCTCCAGCTGGGGCAACAGAGT | 1428 |
| | | | |
| Db | 85 | GAGGCGGAGGTTACAGTGAAGCTGAGATTGTGCCACTGCACCTCCAGCCTGGGTGACAGAGT | 26 |
| QY | 1429 | AAGACTCTGTCTCAAAAAAAAAAAAA 1453 | |
| | | | |
| Db | 25 | GAGATTCCGTCTCAAAAAAAAAAAAA 1 | |

RESULT 13
AA746911
LOCUS
DEFINITION
AA746911 374 bp mRNA EST 22-JAN-1998
nx69c10.s1 NCI-CGAP_Alvl Homo sapiens cDNA clone IMAGE:1267506
similar to contains Alu repetitive element; , mRNA sequence.
ACCESSION
AA746911
NID 92786869
VERSION AA746911.1 GI:2786869
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominini; Homo.
1 (bases 1 to 374)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
On Dec 12, 1995 this sequence version replaced gi:1119179.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:25:55 ; Search time 2910.9 Seconds
(without alignments)
1594.033 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTCAGCGCACACTGTTC.....TCAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_p11:*
8: gb_p12:*
9: gb_p1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_v1:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_cm:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_p1:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 1411.4 | 96.7 | 4740 | 42 | AF135023 |

| | | | | | | |
|----|--------|------|--------|----|------------|----------------------|
| 2 | 1411.4 | 96.7 | 4385 | 42 | AF148532 | AF148532 Homo sapi |
| 3 | 1405 | 96.3 | 5900 | 42 | AF113141 | AF113141 Homo sapi |
| 4 | 296.4 | 20.3 | 1347 | 42 | AF113140 | AF113140 Homo sapi |
| 5 | 248.2 | 17.0 | 166723 | 34 | HS310013 | AL031658 Homo sapi |
| 6 | 240.6 | 16.5 | 151770 | 10 | HSAC002066 | AC002066 Human BAC |
| 7 | 240 | 16.4 | 166663 | 35 | AC007728 | AC007728 Homo sapi |
| 8 | 238.8 | 16.4 | 234053 | 11 | AC002429 | AC002429 Human BAC |
| 9 | 238.6 | 16.4 | 142178 | 35 | AC007160 | AC007160 Homo sapi |
| 10 | 236 | 16.2 | 179757 | 42 | AC006960 | AC006960 Homo sapi |
| 11 | 235 | 16.1 | 134578 | 34 | AC004824 | AC004824 Homo sapi |
| 12 | 235 | 16.1 | 220000 | 35 | AC005674 | AC005674 Homo sapi |
| 13 | 234.6 | 16.1 | 102200 | 9 | HS30M3 | AL031775 Human DNA |
| 14 | 234.6 | 16.1 | 80507 | 11 | AC004024 | AC004024 Homo sapi |
| 15 | 234.6 | 16.1 | 188477 | 34 | AC004971 | AC004971 Homo sapi |
| 16 | 234 | 16.0 | 240000 | 35 | AC006393 | AC006393 Homo sapi |
| 17 | 233.6 | 16.0 | 96625 | 10 | HSAC000118 | AC000118 Human BAC |
| 18 | 233.6 | 16.0 | 135448 | 34 | AC004900 | AC004900 Homo sapi |
| 19 | 232.8 | 16.0 | 39760 | 11 | AC005253 | AC005253 Homo sapi |
| 20 | 232.8 | 16.0 | 46213 | 11 | AC005387 | AC005387 Homo sapi |
| 21 | 232.8 | 16.0 | 42030 | 42 | AC007565 | AC007565 Homo sapi |
| 22 | 232.2 | 15.9 | 45997 | 11 | AC000028 | AC000028 Homo sapi |
| 23 | 232.2 | 15.9 | 45976 | 11 | HS081031 | U81031 Homo sapien |
| 24 | 232.2 | 15.9 | 169026 | 35 | AC007371 | AC007371 Homo sapi |
| 25 | 231.8 | 15.9 | 234596 | 34 | HS94M16 | 297201 Homo sapien |
| 26 | 231.8 | 15.9 | 124956 | 34 | HSJ193N13 | AL078600 Homo sapi |
| 27 | 231.4 | 15.9 | 100000 | 9 | AP000046 | AP000046 Homo sapi |
| 28 | 231.4 | 15.9 | 100000 | 9 | AP000114 | AP000114 Homo sapi |
| 29 | 231.4 | 15.9 | 100000 | 9 | AP000190 | AP000190 Homo sapi |
| 30 | 231.4 | 15.9 | 47559 | 9 | AP000303 | AP000303 Homo sapi |
| 31 | 231.4 | 15.9 | 43012 | 35 | AC007850 | AC007850 Homo sapi |
| 32 | 231.2 | 15.8 | 134970 | 11 | AC002504 | AC002504 Human xp2 |
| 33 | 231.2 | 15.8 | 127587 | 34 | AC005959 | AC005959 Homo sapi |
| 34 | 230.8 | 15.8 | 208643 | 34 | HS44N10 | 297197 Homo sapien |
| 35 | 230.6 | 15.8 | 82359 | 42 | AC004922 | AC004922 Homo sapi |
| 36 | 230.6 | 15.8 | 82359 | 42 | AC004922 | AC004922 Homo sapi |
| 37 | 230.4 | 15.8 | 79376 | 9 | HS454G6 | 298750 Human DNA s |
| 38 | 230.2 | 15.8 | 114784 | 10 | HS1189B24 | AL030996 Human DNA |
| 39 | 230.2 | 15.8 | 164028 | 42 | AC004890 | AC004890 Homo sapi |
| 40 | 230 | 15.8 | 84170 | 9 | HS130N4 | 275887 Human DNA s |
| 41 | 230 | 15.8 | 20733 | 10 | HS356B7 | AL031714 Human DNA s |
| 42 | 230 | 15.8 | 202004 | 11 | H0AF001549 | AF001549 Human Chr |
| 43 | 229.8 | 15.8 | 143747 | 9 | HS370M22 | 282206 Human DNA s |
| 44 | 229.8 | 15.8 | 177236 | 34 | AC005143 | AC005143 Homo sapi |
| 45 | 229.6 | 15.7 | 163871 | 9 | HS1158E12 | AL031584 Human DNA |

ALIGNMENTS

| | | | | | |
|------------|--|------------|-----|-----|-------------|
| RESULT 1 | AF135023 | 4740 bp | DNA | PRI | 20-APR-1999 |
| LOCUS | AF135023 | | | | |
| DEFINITION | Homo sapiens kallikrein-like protein 1 KLK-L1 gene, partial cds. | | | | |
| ACCESSION | AF135023 | | | | |
| NID | g4589272 | | | | |
| VERSION | AF135023.1 | GI:4589272 | | | |
| KEYWORDS | human. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| AUTHORS | Yousef,G.M., Luo,L.Y. and Diamandis,E.P. | | | | |
| TITLE | Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4 | | | | |
| JOURNAL | Biochem. Biophys. Res. Commun. (1999) In press | | | | |
| REFERENCE | 2 (bases 1 to 4740) | | | | |
| AUTHORS | Yousef,G.M., Luo,L.Y. and Diamandis,E.P. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada | | | | |
| FEATURES | Location/Qualifiers | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------|---------|--------------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|----|
| source | 1..4740 | /organism="Homo sapiens" | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | </ |
|--------|---------|--------------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|----|

| | | | |
|---|--|---|------|
| OY | 753 | TTCCACAGCATGGGGCCTGAGGGCGGTGACCTCCACCATAAGAAAATCCTTATAAC | 812 |
| Db | 3562 | TTCCACAGCATGGGGCCTGAGGGCGGTGACCTCCACCATAAGAAAATCCTTATAAC | 3621 |
| OY | 813 | TTTTGACTCCCCAAAAAACCCTGACTAGAAAATAGCCTTACTGTTGACGGGAGCCTTACCAAT | 872 |
| Db | 3622 | TTTTGACTCCCCAAAAAACCCTGACTAGAAAATAGCCTTACTGTTGACGGGAGCCTTACCAAT | 3681 |
| OY | 873 | AACATAAATAGTCGATTTATGCATACGTTTATGCATTCATGATATACCTTTGTTGAAT | 932 |
| Db | 3682 | AACATAAATAGTCGATTTATGCATACGTTTATGCATTCATGATATACCTTTGTTGAAT | 3741 |
| OY | 933 | TTTTGATATTTCTAAGCTACACAGTTCGCTGTGTAATTTTAAATTTGTTGCAACTCT | 992 |
| Db | 3742 | TTTTGATATTTCTAAGCTACACAGTTCGCTGTGTAATTTTAAATTTGTTGCAACTCT | 3801 |
| OY | 993 | CCTAAATTTTCTGATGTGTTTATTTGAAAAATCCAGTATAAGTGACTTGTGCA-TT | 1051 |
| Db | 3802 | CCTAAATTTTCTGATGTGTTTATTTGAAAAATCCAGTATAAGTGACTTGTGCAAGTT | 3861 |
| OY | 1052 | CAAAACAGGGTTGTTCAAGGGTCAACTGTGTACCCAGAGGGAACAGTGACACAGATTCA | 1111 |
| Db | 3862 | CAAAACAGGGTTGTTCAAGGGTCAACTGTGTACCCAGAGGGAACAGTGACACAGATTCA | 3921 |
| OY | 1112 | TAGAGGTGAAACACAGAGAGAAACAGGAAAAATCAAGACTCTACAAGAGGCTGGCAGG | 1171 |
| Db | 3922 | TAGAGGTGAAACACAGAGAGAAACAGGAAAAATCAAGACTCTACAAGAGGCTGGCAGG | 3981 |
| OY | 1172 | GTCGCTCATGCTGTAAATCCACAGCACTTTGGGAGCGGAGGCAGGCAATCACTTGAGTA | 1231 |
| Db | 3982 | GTCGCTCATGCTGTAAATCCACAGCACTTTGGGAGCGGAGGCAGGCAATCACTTGAGTA | 4041 |
| OY | 1232 | AGGAGTTCAAGACCAGCCTGGCCAAAATGTTGAATCCTGTCTGTACTTAAATAATACAAA | 1291 |
| Db | 4042 | AGGAGTTCAAGACCAGCCTGGCCAAAATGTTGAATCCTGTCTGTACTTAAATAATACAAA | 4101 |
| OY | 1292 | GTTAGCTGATATGTTGTCAGGGCGCCTGTAAATCCAGCTACTTTGGAGGCTGAGGCAGGA | 1351 |
| Db | 4102 | GTTAGCTGATATGTTGTCAGGGCGCCTGTAAATCCAGCTACTTTGGAGGCTGAGGCAGGA | 4161 |
| OY | 1352 | GAATTCCTTGAATATGGGAGGCGAGAGGTTGAAGTGAAGTTGAGATCAACCACTATACTCC | 1411 |
| Db | 4162 | GAATTCCTTGAATATGGGAGGCGAGAGGTTGAAGTGAAGTTGAGATCAACCACTATACTCC | 4221 |
| OY | 1412 | AGCTGGGGCAACAGAGTAAGACTCTGTCTCAAAAAAATAAAAAA | 1459 |
| Db | 4222 | AGCTGGGGCAACAGAGTAAGACTCTGTCTCAAAAAAATAAAAAA | 4269 |
| RESULT 2 | | | |
| AF148532 | | AF148532 | |
| LOCUS | AF148532 | 4385 bp | DNA |
| DEFINITION | Homo sapiens kallikrein 4 (KLK4) gene, complete cds. | | |
| ACCESSION | AF148532 | | |
| NID | 95020095 | | |
| VERSION | AF148532.1 GI:5020095 | | |
| KEYWORDS | | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | |
| Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | |
| REFERENCE | 1 (bases 1 to 4385) | | |
| AUTHORS | Stephenson,S.A., Verity,K., Ashworth,L. and Clements,J.A. | | |
| TITLE | Localization of a new prostate specific antigen-related serine | | |
| protease gene, KLK4, is evidence for an expanded human kallikrein | | | |
| (KLK) gene family cluster on chromosome 19q13.3-13.4 | | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 4385) | | |
| AUTHORS | Stephenson,S.A., Ashworth,L. and Clements,J.A. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (03-MAY-1999) School of Life Science, Queensland | | |
| University of Technology, George Street, Brisbane, Queensland 4001. | | | |


```
FEATURES             Australia
source               1..4385
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="19"
                    /map="19q13.3-q13.4"
                    join(<551..611,1875..2037,2459..2709,2793..2929,
                    4200..>4352)
mRNA
                    /gene="KLK4"
                    /product="kallikrein 4"
                    <551..>4352
gene
                    /gene="KLK4"
                    join(551..611,1875..2037,2459..2709,2793..2929,4200..4352)
CDS
                    /gene="KLK4"
                    /function="serine protease"
                    /codon_start=1
                    /product="kallikrein 4"
                    /protein_id="AAD38019.1"
                    /db_xref="PID:g5020096"
                    /db_xref="GI:5020096"
                    /translation="MATAGNPMWGFLLGLIVAGSLVSGSQIINGEDCSPHSQPW
                    QALVMENELFCSGVLVHPQWYLSAHCFONSTYIGLHLSLEADQEPGSQMEASLS
                    VRHPEYNRPFLANDLMLIKLIDESVSDTIRISISASOCPTAGNSCLVSGWGLANGR
                    MPTVLQCVNVSVEVSEVSKLYDPLHYPSMFCAGGQDQKDSNGDSGGLICNGLYLQ
                    GLVSEGKAPCGGVGPVGYVTNLCKFTEWIEKTVQAS"

BASE COUNT      1013 a      1184 c      1192 g      996 t
ORIGIN
Query Match      96.7%; Score 1411.4; DB 42; Length 4385;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY      33  CAGAGCTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGG  92
      || |||||
Db      2454 CACAGCTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGG  2513

QY      93  AGCCAGATGGTGAGGCCAGCCTCTCCGTACGGGCACCCAGAGTACAACAGACCCTGCTC  152
      |||||
Db      2514 AGCCAGATGGTGAGGCCAGCCTCTCCGTACGGGCACCCAGAGTACAACAGACCCTGCTC  2573

QY      153 GCTACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGG  212
      |||||
Db      2574 GCTACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGG  2633

QY      213 AGCATCAGCATTGTCTGCGAGTGCCTTACCGCGGGGAACCTTTGCTTCTGGCTGG  272
      |||||
Db      2634 AGCATCAGCATTGTCTGCGAGTGCCTTACCGCGGGGAACCTTTGCTTCTGGCTGG  2693

QY      273 GGTCTGTGGCGAAGGTGAGCTACGGGCTGTGTCTGCCCCCTCTTCAAGAGAGTCTCT  332
      |||||
Db      2694 GGTCTGTGGCGAAGGTGAGCTACGGGCTGTGTCTGCCCCCTCTTCAAGAGAGTCTCT  2753

QY      333 GCCCAGTCGCGGGGCTGACCCAGAGCTTCGCTCCACGAGCAATGCTTACCCTGCTGC  392
      |||||
Db      2754 GCCCAGTCGCGGGGCTGACCCAGAGCTTCGCTCCACGAGCAATGCTTACCCTGCTGC  2813

QY      393 AGTGCCTGAACGTTCGGTGTCTTGANAGGCTCTGCANTAGCTCTATGACCCGCTGT  452
      |||||
Db      2814 AGTGCCTGAACGTTCGGTGTCTTGAGGAGGCTCTGCAGTAAGCTCTATGACCCGCTGT  2873

QY      453 ACCACCCCAACATGTTCTGCGCGCGGAGGGCAAGACAGACAGAGACTCTGCAACGTGA  512
      |||||
Db      2874 ACCACCCCAACATGTTCTGCGCGCGGAGGGCAAGACAGAGAGACTCTGCAACGTGA  2933

QY      513 GAGAGGGGAAAGGGGAGGCGAGCTCAGGGAAGGCTGAGAGAGGGGAGACAGAGAC  572
      |||||
Db      2934 GAGAGGGGAAAGGGGAGGCGAGCTCAGGGAAGGCTGAGAGAGGGGAGACAGAGAC  2993

QY      573 ACACAGGGCCGATGGCGAGATGACAGAGATGAGAGACACACAGAGGAGACAGTGAACA  632
      |||||
Db      2994 ACACAGGGCCGATGGCGAGATGACAGAGATGAGAGACACACAGAGGAGACAGTGAACA  3053
```

```
QY      633  AGAGAGAGAACTGAGAGAAACAGAGAAATTAACACAGCAATTAAGAGAGCAAAAGGAG  692
      |||||
Db      3054  AGAGAGAGAACTGAGAGAAACAGAGAAATTAACACAGCAATTAAGAGAGCAAAAGGAG  3113

QY      693  AGAGAAACAGAAACAGACATGGGGAGGCGAGAAACACACACATAGAAATGCAAGTTGACC  752
      |||||
Db      3114  AGAGAAACAGAAACAGACATGGGGAGGCGAGAAACACACACATAGAAATGCAAGTTGACC  3173

QY      753  TTCCAACAGCATGGGGCTGAGGGCGGTGACCTCCACCCAATAGAAATCCTTTATAAC  812
      |||||
Db      3174  TTCCAACAGCATGGGGCTGAGGGCGGTGACCTCCACCCAATAGAAATCCTTTATAAC  3233

QY      813  TTTGACTCCCAAAACCTGACTAGAAATAGCCTACTGTGACGGGAGCCTTACCAAT  872
      |||||
Db      3234  TTTGACTCCCAAAACCTGACTAGAAATAGCCTACTGTGACGGGAGCCTTACCAAT  3293

QY      873  AACATAATATAGTCGATTATGCAATCGTTTATGCATTCATATATACCTTTGTTGAAT  932
      |||||
Db      3294  AACATAATATAGTCGATTATGCAATCGTTTATGCATTCATATATACCTTTGTTGAAT  3353

QY      933  TTTTGATATTTCTAAGCTACACAGTTCGTCTGTGAATTTTAAATTGTTGCAACTCT  992
      |||||
Db      3354  TTTTGATATTTCTAAGCTACACAGTTCGTCTGTGAATTTTAAATTGTTGCAACTCT  3413

QY      993  CCTAAATTTTCTGATGTGTTTATTGAAAAATCCAAGTATAAGTGACTGTGCA-TT  1051
      |||||
Db      3414  CCTAAATTTTCTGATGTGTTTATTGAAAAATCCAAGTATAAGTGACTGTGCAAGTT  3473

QY      1052 CAACCAAGGTTGTTCAAGGGTCAACTGTGTACCACAGGGAACACAGTGACACAGATTCA  1111
      |||||
Db      3474  CAACCAAGGTTGTTCAAGGGTCAACTGTGTACCACAGGGAACACAGTGACACAGATTCA  3533

QY      1112 TAGAGGTGAACACGAGAGAGAAACAGAAAAATCAAGACTTACAAAGAGGCTGGGCAGG  1171
      |||||
Db      3534  TAGAGGTGAACACGAGAGAGAAACAGAAAAATCAAGACTTACAAAGAGGCTGGGCAGG  3593

QY      1172 GTGGCTCATGCTGTAAATCCACGACTTTGGGAGGCGAGGCGAGCAGATCACTTGAGGTA  1231
      |||||
Db      3594  GTGGCTCATGCTGTAAATCCACGACTTTGGGAGGCGAGGCGAGCAGATCACTTGAGGTA  3653

QY      1232 AGGAGTTCAAGACCAAGCCTGGGCCAAATGTTGAATCCTGTCTACTAAATAACAATA  1291
      |||||
Db      3654  AGGAGTTCAAGACCAAGCCTGGGCCAAATGTTGAATCCTGTCTACTAAATAACAATA  3713

QY      1292 GTTAGCTGATATGTTGGCAGGCGCCTGTATATCCAGCTACTTGGAGGCTGAGGCAGGA  1351
      |||||
Db      3714  GTTAGCTGATATGTTGGCAGGCGCCTGTATATCCAGCTACTTGGAGGCTGAGGCAGGA  3773

QY      1352 GAATTGCTGAATATGAGGAGGAGAGGTTGAAGTGAGTTGAGATCACACCACTATACTCC  1411
      |||||
Db      3774  GAATTGCTGAATATGAGGAGGAGAGGTTGAAGTGAGTTGAGATCACACCACTATACTCC  3833

QY      1412 AGCTGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA  1459
      |||||
Db      3834  AGCTGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA  3881

RESULT      3
AF113141
LOCUS      AF113141      5900 bp      DNA      PRI      25-MAR-1999
DEFINITION Homo sapiens serine protease prostate gene, complete cds.
ACCESSION  AF113141
NID      94512031
VERSION  AF113141.1  GI:4512031
KEYWORDS
SOURCE
ORGANISM      human.
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 5900)
AUTHORS      Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
              wang,K.
```


TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 5900)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA
FEATURES
source location/Qualifiers
1. 5900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
join(<1411. 1471,2735. 2897,3319. 3569,3653. 3789, 5062. 5817)
/product="serine protease prostase"
1411. 1471
/number=1
join(1411. 1471,2735. 2897,3319. 3569,3653. 3789, 5062. 5214)
/note="androgen regulated"
/codon_start=1
/product="serine protease prostase"
/protein_id="AAD21581.1"
/db_xref="PID:g4512032"
/db_xref="GI:4512032"
/translation="MATAGNPMWGFILYLILGVAGSLVSGSCQIINGEDCSPHSQPM QALVMENELFCGVLVHPQWVLSAHCFOINSYITIGLHSLLEADQEPGSOMVEASLS VRHPEYNRLPLANDMLIKLIDESYSDTIRISISIASQPTAGNSCLVSGWGLIANGR MPTVLCVNVSVSEVCSKLYDPLYHPSMFCAGGHDQDSCNDSGGPLICNGYLQ GLVSFGKAPCGQVGPVYTNLCKFTIEWIEKTVQAS"
exon 2735. 2897
/number=2
exon 3319. 3569
/number=3
exon 3653. 3789
/number=4
exon 5062. 5817
/number=5
BASE COUNT 1417 a 1643 c 1569 g 1271 t
ORIGIN
Query Match 96.3%; Score 1405; DB 42; Length 5900;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1419; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 33 CAGAGCTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGCCAGGG 92
Db 3314 CACAGCTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGCCAGGG 3373
QY 93 AGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGGCACCCAGAGTACAACAGACCCTGCTC 152
Db 3374 AGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGGCACCCAGAGTACAACAGACCCTGCTC 3433
QY 153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGG 212
Db 3434 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGG 3493
QY 213 AGCATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAACTCTTGCTGTTCTGGCTGG 272
Db 3494 AGCATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAACTCTTGCTGTTCTGGCTGG 3553
QY 273 GGTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTCTCAAGAGGTCTCT 332
Db 3554 GGTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTCTCAAGAGGTCTCT 3613
QY 333 GCCCAGTCGGGGGGGTGACCCAGAGCTGTGCTCCAGGCAAGATGCCCTACCGTGTGC 392
Db 3333 GCCCAGTCGGGGGGGTGACCCAGAGCTGTGCTCCAGGCAAGATGCCCTACCGTGTGC 392

Db 3614 GCCCAGTCGGGGGGGTGACCCAGAGCTGTGCTCCAGGCAAGATGCCCTACCGTGTGC 3673
QY 393 AGTGGTGAACGTGTCCGGTGTGTCTGANGAGGTCTGCANTAGCTCTATGACCCGCTGT 452
Db 3674 AGTGGTGAACGTGTCCGGTGTGTCTGANGAGGTCTGCAGTAGCTCTATGACCCGCTGT 3733
QY 453 ACCACCCANCATGTTCTGCGCGCGCGGCGGCAAGACCAGAGAGACTCTCGCAACGTGA 512
Db 3734 ACCACCCAGCATGTTCTGCGCGCGCGGCGGCAAGACCAGAGAGACTCTCGCAACGTGA 3793
QY 513 GAGAGGGGAAAGGGGAGGGCCAGGCGACTCAGGGAAGGGTGGAGAAAGGGGAGACAGAGAC 572
Db 3794 GAGAGGGGAAAGGGGAGGGCCAGGCGACTCAGGGAAGGGTGGAGAAAGGGGAGACAGAGAC 3853
QY 573 ACACAGGGCCGATGGCGAGATGCAGAGATGGAGAGACACACAGGAGACAGTGAACA 632
Db 3854 ACACAGGGCCGATGGCGAGATGCAGAGATGGAGAGACACACAGGAGACAGTGAACA 3913
QY 633 AGAGAGAAACTGAGAGAAACAGAGAAATAAACACAGAAATAAAGAGCAAGCAAGAG 692
Db 3914 AGAGAGAAACTGAGAGAAACAGAGAAATAAACACAGAAATAAAGAGCAAGCAAGAG 3973
QY 693 AGAGAAACAGAAACAGACATGGGGAGGCGAAGAACACACACACATAGAAATGCAGTTGACC 752
Db 3974 AGAGAAACAGAAACAGACATGGGGAGGCGAAGAACACACACACATAGAAATGCAGTTGACC 4033
QY 753 TTCCAACAGCATGGGGCCTGAGGGCGGTGACCTCCACCACCAATAGAAAATCCTCTTATTAAC 812
Db 4034 TTCCAACAGCATGGGGCCTGAGGGCGGTGACCTCCACCACCAAGAAAATCCTCTTATTAAC 4093
QY 813 TTTTGACTCCCCAAAAACCTGACTAGAAATAGCCTACTGTGTGACGGGGAGCCTTACCAAT 872
Db 4094 TTTTGACTCCCCAAAAACCTGACTAGAAATAGCCTACTGTGTGACGGGGAGCCTTACCAAT 4153
QY 873 AACATTAATAGTCGATTTATGATACGTTTATGATTCATCATGATATACCTTTGTTGAAT 932
Db 4154 AACATTAATAGTCGATTTATGATACGTTTATGATTCATCATGATATACCTTTGTTGAAT 4213
QY 933 TTTTGATATTTCTAGCTACACAGTTCGCTGTGAATTTTAAATTGTTGCAACTCT 992
Db 4214 TTTTGATATTTCTAGCTACACAGTTCGCTGTGAATTTTAAATTGTTGCAACTCT 4273
QY 993 CCTAAATTTTCTGATGTGTTTATGAAAAAATCCAAGTATAGTGGACTGTGCA-TT 1051
Db 4274 CCTAAATTTTCTGATGTGTTTATGAAAAAATCCAAGTATAGTGGACTGTGCACTT 4333
QY 1052 CAAACCAAGGTTGTCAAGGGTCAACTGTGTACCCAGAGGGAACAGTGACACAGATTCA 1111
Db 4334 CAAACCAAGGTTGTGTCAAGGGTCAACTGTGTACCCAGAGGGAACAGTGACACAGATTCA 4393
QY 1112 TAGAGGTGAACACGAGAGAGAAACAGAAAAATCAAGACTCTACAAAAGAGGCTGGGCAAG 1171
Db 4394 TAGAGGTGAACACGAGAGAGAAACAGAAAAATCAAGACTCTACAAAAGAGGCTGGGCAAG 4453
QY 1172 GTGGCTCATGCTGTATATCCAGCACTTTGGGAGGGGAGGCGAGAGATCACTTGAGGTA 1231
Db 4454 GTGGCTCATGCTGTATATCCAGCACTTTGGGAGGGGAGGCGAGAGATCACTTGAGGTA 4513
QY 1232 AGGAGTTCAAGACCAAGCCTGGCCAAATGTGAAATCCTGTCTGTACTAAAAATACAAA 1291
Db 4514 AGGAGTTCAAGACCAAGCCTGGCCAAATGTGAAATCCTGTCTGTACTAAAAATACAAA 4573
QY 1292 GTTAGCTGATATGTGGCAGGGCGCTGTATCCAGCTACTTGGGAGGCTGAGGCAAG 1351
Db 4574 GTTAGCTGATATGTGGCAGGGCGCTGTATCCAGCTACTTGGGAGGCTGAGGCAAG 4633
QY 1352 GAATGCTTGAATATGGAGGAGAGGTTGAGTGAAGTGAATGAGATCACACCACTATACTCC 1411
Db 4634 GAATGCTTGAATATGGAGGAGAGGTTGAGTGAAGTGAATGAGATCACACCACTATACTCC 4693
QY 1412 AGCTGGGGCAACAGAGTAAGACTGTGTCAAAAAAAAAAAAAAAAAAAAA 1459
Db 4694 AGCTGGGGCAACAGAGTAAGACTGTGTCAAAAAAAAAAAAAAAAAAAAA 4741


```
RESULT      4  
LOCUS       AF113140             1347 bp    mRNA          PRI           25-MAR-1999  
DEFINITION Homo sapiens serine protease prostate mRNA, complete cds.  
ACCESSION   AF113140  
NID         g4512029  
VERSION     AF113140.1 GI:4512029  
KEYWORDS    human.  
SOURCE      Homo sapiens  
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
            Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE   1 (bases 1 to 1347)  
AUTHORS     Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.  
TITLE        Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression  
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)  
MEDLINE     99179024  
REFERENCES   2 (bases 1 to 1347)  
AUTHORS     Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.  
TITLE        Direct Submission  
JOURNAL      Submitted (10-DEC-1998) Chirosence R & D, 1631 220th St SE,  
Bothell, WA 98021, USA  
FEATURES  
source       Location/Qualifiers  
              1..1347  
               /organism="Homo sapiens"  
               /db_xref="taxon:9606"  
               /chromosome="19"  
               /map="19q13"  
               1..765  
                /note="androgen regulated"  
                /codon_start=1  
                /product="serine protease prostate"  
                /protein_id="AAD21580.1"  
                /db_xref="PID:g4512030"  
                /db_xref="GI:4512030"  
                /translation="MATAGNPWGWFLGYLLGVAGSLVSGSCSOIINGEDCSPHSOPM QALVMENELFCGVLVPQWLSAHGFGNSYITGLHLSEADPEGSOMVEASLS VRHPYNRPLANDLMLIKLDSESSEDTRISIASOCPTAGNSCLYSWGMLLANGR MPYLQCVNVSVSEEVC SKLYDPLYHPSMECAGGGHDQKDSNGSGGPLICNGYLQ GLVSFGRAPCGOVGPVYTNLCKFTENELEKTIVQA"  
polyA_site   1327..1332  
BASE COUNT   269 a      489 c      334 g      255 t  
ORIGIN  
Query Match  20.3%; Score 296.4; DB 42; Length 1347;  
Best Local Similarity 81.0%; Pred.No. 2e-59;  
Matches 393; Conservative 0; Mismatches 9; Indels 83; Gaps 1;
```

```

Db      452 CTGGCTGGGGTCTGCTGGCGAAC----- 475
QY      325 GGTCCTCTGCCAGTCGGCGGGGCTGACCAGAGCTCTGCGTCCAGGCAGATGCTTAC 384
Db      475 -----GGCAGAAATGCCCTAC 488
QY      385 CGTGCTGCAGTGCCTGAACGCTGTCCGGTGTCTGANGAGGCTCTGCANTAAAGCTCTATGA 444
Db      489 CGTGCTGCAGTGCCTGAACGCTGTCCGGTGTCTGAGAGAGTCTGCAGTAAGCTCTATGA 548
QY      445 CCCGCTGTACCAACCCCANCAATGTTCTGCGCCGCGGAGGCGCAAGACCAAGAGACTCTG 504
Db      549 CCCGCTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGCGACCAAGAGACTCTG 608
QY      505 CAACG 509
Db      609 CAACG 613

RESULT 5
HS310013 LOCUS HS310013 166723 bp DNA HTG 10-JUN-1999
DEFINITION Homo sapiens chromosome 20 clone 310013, WORKING DRAFT SEQUENCE, in
ACCESSION AL031658
NID 94902630
VERSION AL031658.10 GI:4902630
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166723)
AUTHORS Kay, M.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hunquerry@anger.ac.uk Clone requests: clonerequests@anger.ac.uk
On May 27, 1999 this sequence version replaced gi:4741461.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj310013 Contig_ID: 00252 acc=AL031658
length: 1113 bp Unfinished: dj310013 Contig_ID: 00912
acc=AL031658 length: 162976 bp Unfinished: dj310013 Contig_ID:
02100 acc=AL031658 length: 1034 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1.166723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="310013"

BASE COUNT 42104 a 40443 c 39486 g 43089 t 1601 others
ORIGIN
Query Match 17.0%; Score 248.2; DB 34; Length 166723;
Best Local Similarity 84.3%; Pred. No. 4.8e-48;
Matches 291; Conservative 0; Mismatches 53; Indels 1; Gaps 1;

QY 1113 AGAGGTGAACACGAGAGAAACAGGAAAAATCAAGACTCTACAAAGAGGCTGGCGAGGG 1172
Db 93064 AAAATTTCAACATCTGTAAAAAAGATAGACTTCAAAAACATCGCTAAGTTGCTGGGCATGG 93123
QY 1173 TGGCTCATGCTGTAAATCCAGCACTTTGGAGG-CGAGGCAGGCAGATCACTGAGGTA 1231

```


Db 93124 TGGCTCATGCTGTAAATCCAGCACTTTGGGAGCCGAGGTGGAGATCACTTGAGGTC 93183
QY 1232 AGGAGTTCAAGACCAAGCCTGGCCAAATGTTGAATCCTGTCTGTACTAAATACAAA 1291
Db 93184 AGGAGTTCAAGACCAAGCCTGGCCAAATGTTGAATCCTGTCTGTACTAAATACAAA 93243
QY 1292 GTTAGCTGATATGTTGGCAGGCGCCTGTAAATCCCACTACTTGGAGGCTTGAGCAGCA 1351
Db 93244 ATTAGCTGGGCATGTGGCACACAGCCCTGTAGTCCCACTACTCAGGAGGCTTGAGGCA 93303
QY 1352 GAATTGCTGAATATGGGAGGCGAGAGGTTGAAGTGAGTTGAGATCACACCCTATACTCC 1411
Db 93304 GAATTGCTGAACCTGGGAGGCGGAGGTTGCACTGAGCTGAGATCACACCCTGTACTCC 93363
QY 1412 AGCTGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAA 1456
Db 93364 AGCCTGGGCGACAGAGTGAAGACTCTGTCTCAAAAAAAAAAAAAA 93408

RESULT 6
LOCUS HSAC002066 151770 bp DNA PRI 09-MAY-1997
DEFINITION Human BAC clone RG030H15 from 7q31, complete sequence.
ACCESSION AC002066
NID 92076721
VERSION AC002066.1 GI:2076721
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 151770)
Dante,M, Wamsley,P and Keppler,D.
The sequence of H. sapiens BAC clone RG030H15
Unpublished (1997)
2 (bases 1 to 151770)
Waterston,R.
Direct Submission
Submitted (09-MAY-1997)
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
http://genome.wustl.edu/gsc
e-mail: saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/DTB/CHR7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:
This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
VECTOR: PBEL0
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of H_RG030H15;
actual end is at 151770 of H_RG030H15. The orientation of this clone is unknown.

This clone contains SRS SWSS965.
FEATURES
source location/Qualifiers
1. .151770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RG030H15"
/clone_lib="CITB-978SK-B"
/map="7q31"
repeat_region complement(671. .690)
/rpt_family="L1"
repeat_region 2544. .2593
/rpt_family="L1"
repeat_region complement(2759. .2781)
/rpt_family="L1"
repeat_region complement(4347. .4468)
/rpt_family="L1"
repeat_region complement(5317. .5615)
/rpt_family="ALU"
repeat_region complement(6063. .6314)
/rpt_family="ALU"
repeat_region complement(7577. .8152)
/rpt_family="L1"
repeat_region complement(8176. .8233)
/rpt_family="L1"
repeat_region complement(8706. .9002)
/rpt_family="ALU"
repeat_region complement(9318. .9390)
/rpt_family="ALU"
repeat_region complement(9403. .9434)
/rpt_family="L1"
repeat_region complement(9481. .9510)
/rpt_family="L1"
repeat_region 9513. .9789
/rpt_family="ALU"
repeat_region 10325. .10770
/rpt_family="L1"
misc_feature 12726. .13009
/note="similar to human EST D45516 (NID:g662470)"
repeat_region 14161. .15579
/rpt_family="L1"
repeat_region 15804. .15902
/rpt_family="L1"
repeat_region 17101. .17218
/rpt_family="ALU"
repeat_region complement(17462. .17485)
/rpt_family="L1"
repeat_region complement(17519. .17675)
/rpt_family="ALU"
repeat_region complement(18866. .19055)
/rpt_family="L1"
repeat_region complement(19068. .20712)
/rpt_family="L1"
repeat_region complement(20933. .21001)
/rpt_family="L1"
repeat_region complement(21021. .21785)
/rpt_family="L1"
repeat_region complement(21919. .21943)
/rpt_family="L1"
repeat_region 23706. .24283
/rpt_family="L1"
repeat_region 25651. .25942
/rpt_family="ALU"
repeat_region 27166. .27287
/rpt_family="ALU"
repeat_region 27292. .27794
/rpt_family="L1"
repeat_region complement(27974. .28348)

| | | |
|---------------|--|--|
| repeat_region | /rpt_family="L1" 28315. .28348 | |
| repeat_region | /rpt_family="L1" complement(30214. .30233) | |
| repeat_region | /rpt_family="L1" complement(30338. .30564) | |
| repeat_region | /rpt_family="L1" 30584. .30882 | |
| repeat_region | /rpt_family="ALU" 30894. .30920 | |
| repeat_region | /rpt_family="L1" complement(30906. .31081) | |
| repeat_region | /rpt_family="L1" complement(31082. .31447) | |
| repeat_region | /rpt_family="ALU" complement(31833. .32074) | |
| repeat_region | /rpt_family="L1" complement(32115. .32165) | |
| repeat_region | /rpt_family="L1" complement(32333. .33181) | |
| repeat_region | /rpt_family="L1" 34263. .38923 | |
| repeat_region | /note="Difuse retroviral homology. See Z76735 (NID:g1438501)." | |
| repeat_region | complement(34263. .34419) | |
| repeat_region | /note="similar to human EST AA151971 (NID:g1720827) zo30b03.r1" | |
| repeat_region | complement(34263. .34489) | |
| repeat_region | /note="similar to human EST C18895 (NID:g1580497) " complement(34263. .34526) | |
| repeat_region | /note="similar to human EST AA133084 (NID:g1694609) zo22h03.r1" | |
| repeat_region | complement(34474. .34547) | |
| repeat_region | /note="similar to human EST AA132265 (NID:g1693753) zo27a11.r1" | |
| repeat_region | complement(34549. .34671) | |
| repeat_region | /note="similar to human EST AA055887 (NID:g1548225) zl76h06.r1" | |
| repeat_region | complement(34549. .34700) | |
| repeat_region | /note="similar to human EST AA132265 (NID:g1693753) zo27a11.r1" | |
| repeat_region | complement(34549. .34666) | |
| repeat_region | /note="similar to human EST AA134504 (NID:g1692069) zo12a05.r1" | |
| repeat_region | complement(34659. .34775) | |
| repeat_region | /note="similar to human EST AA152341 (NID:g1719175) zo01g11.r1" | |
| repeat_region | complement(34762. .35004) | |
| repeat_region | /note="similar to human EST C18935 (NID:g1580537) " complement(34881. .34991) | |
| repeat_region | /note="similar to human EST C04699 (NID:g1467950) " complement(35024. .35255) | |
| repeat_region | /note="similar to human EST AA099408 (NID:g1645301) zl77a05.s1" | |
| repeat_region | complement(35025. .35255) | |
| repeat_region | /note="similar to human EST T51824 (NID:g653684) " yb54b04.s1" | |
| repeat_region | 35321. .35361 | |
| repeat_region | /note="similar to human EST AA160337 (NID:g1734916) zo56e05.s1" | |
| repeat_region | 35361. .35669 | |
| repeat_region | /note="similar to human EST T03007 (NID:g457094) " complement(35694. .36004) | |
| repeat_region | /note="similar to human EST AA135325 (NID:g1697193) zo28e05.r1" | |
| repeat_region | complement(36078. .36282) | |
| repeat_region | /note="similar to human EST AA152341 (NID:g1719175) zo01g11.r1" | |
| repeat_region | complement(36078. .36414) | |
| repeat_region | /note="similar to human EST AA149780 (NID:g1720723) zo01a08.r1" | |
| repeat_region | 36141. .36342 | |
| repeat_region | /note="similar to human EST T61055 (NID:g664092) | |

[illegible]


```

AC007728/c AC007728 166663 bp DNA HTG 05-JUN-1999.
DEFINITION Homo sapiens chromosome 16 clone 327F22, WORKING DRAFT SEQUENCE, 5
ordered pieces.
ACCESSION AC007728
NID 95001450
VERSION AC007728.1 GI:5001450
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166663)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
Han,C. and Deaven,L.
TITLE Sequencing of Human Chromosome 16q12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166663)
AUTHORS Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 166663)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 22988: contig of 22988 bp in length
* gap of unknown length
* 22989 109379: contig of 86391 bp in length
* gap of unknown length
* 109380 154579: contig of 45200 bp in length
* gap of unknown length
* 154580 156226: contig of 1647 bp in length
* gap of unknown length
* 156227 166663: contig of 10437 bp in length.
* Location/Qualifiers
* 1. 166663
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone="327F22"
* /chromosome="16"
BASE COUNT 51110 a 37036 c 34518 g 43999 t
ORIGIN
Query Match 16.4%; Score 240; DB 35; Length 166663;
Best Local Similarity 83.5%; Pred. No. 3.9e-46;
Matches 284; Conservative 0; Mismatches 55; Indels 1; Gaps 1
QY 1121 AACACGAAGAGAAACAGGAAAAATCAAGACTTACAAAGAGCGTGGCGAGGCTGCAT 1180
1 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 111668 AGCAGCACCAACACACCATCAACAGAAAGACACAAATAAGCGCGGCACATTGGCTCAT 111609
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 1181 GCCTGTATCCAGCAGCTTTGGGAGGC-GAGGCAGGCAGATCACTTGAGTGAAGAGATTG 1239
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

Db 111608 GCCTGTAATCCACGACACTTTGGGAGGCGTACGCGAGCAGATCACTTGAGGTCAGAGTTT 111549
QY 1240 AAGACCAGCCTGGCCAAATGGTGAATCCTGTCTACTATAAAATACAAAGTTAGCTG 1299
      |||||
Db 111548 GAGACCAGCCTGGTCAATATGGTGAACCCTGTGTCTACTAAAAATACAAAAATTAGCTG 111489
QY 1300 GATATGTTGGCAGCGCGCTGTATATCCAGCTACTTGGGAGGCTGAGGAGGAGAAATTGCT 1359
      |||||
Db 111488 GGTGTGGTGGCAGGCACTGTATATCCAGCTACTTGGGAGGCTGAGGAGGAAATTGCT 111429
QY 1360 TGAATATGGAGGCGAGAGGTTGAAGTGAAGTGAATCACCACCTATATCCAGCTGGGG 1419
      |||||
Db 111428 TGAACCTGGAGCGCGAGGTTGCAGTGAAGTGAATGCACCATTTGCACCTCCAGCCTAGG 111369
QY 1420 CAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
      |||||
Db 111368 CAACAGAGCAAGACTCTGTCTCAAAACAAAGAAACAACA 111329

RESULT 8
AC002429 234053 bp DNA PRI 19-AUG-1997
LOCUS Human BAC clone GS200K05 from 7q21-q22, complete sequence.
DEFINITION AC002429
ACCESSION 92335067
NID AC002429.1 GI:2335067
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homini; Homo.
REFERENCE 1 (bases 1 to 234053)
AUTHORS Graves,T, Duckels,G and Hawkins,M.
TITLE The sequence of H. sapiens BAC clone GS200K05
JOURNAL Unpublished (1997)
AUTHORS 2 (bases 1 to 234053)
REFERENCE Waterston,R.
JOURNAL Direct Submission
AUTHORS Submitted (19-AUG-1997) Department of Genetics, Washington
REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE SUBMITTED BY:
JOURNAL Genome Sequencing Center
AUTHORS Department of Genetics
REFERENCE Washington University
AUTHORS St. Louis MO 63108, USA
REFERENCE http://genome.wustl.edu/gsc
JOURNAL mailto:sapiens@watson.wustl.edu

COMMENT
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 or send
mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc.
(http://www.genomesystems.com).

```


Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelOBAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of 1 GS200K05;
actual end is at 234053 of GS200K05. The orientation of this clone
is unknown.

This cosmid contains sts's SWS53537 (NID:g1222821) and SWS53746
(NID:g1916514).

FEATURES

source
1. .234053
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="GS200K05"
/clone_1b="GSBAC1"
/map="7q21-q22"
complement(47. .74)
/rpt_family="L1"
complement(145. .171)
/rpt_family="L1"
1042. .1102
/rpt_family="L1"
complement(3234. .3271)
/rpt_family="L1"
complement(3293. .3329)
/rpt_family="L1"
complement(3345. .3560)
/rpt_family="L1"
complement(4040. .7521)
/rpt_family="L1"
5186. .5603
/rpt_family="L1"
5764. .5791
/rpt_family="L1"
complement(7522. .7598)
/rpt_family="ALU"
complement(7599. .8058)
/rpt_family="L1"
complement(8059. .8335)
/rpt_family="ALU"
complement(8336. .10510)
/rpt_family="L1"
complement(10522. .10654)
/rpt_family="ALU"
10731. .10962
/note="match to EST AA011406 (NID:g1472432) z101g01.s1"
complement(11288. .11342)
/rpt_family="L1"
complement(12291. .12332)
/rpt_family="L1"
12823. .12860
/rpt_family="L1"
complement(13091. .13119)
/rpt_family="L1"
complement(13605. .13894)
/rpt_family="ALU"
14412. .14433
/rpt_family="L1"
complement(16228. .16474)
/rpt_family="ALU"
17447. .17474
/rpt_family="L1"
complement(17649. .17938)
/rpt_family="ALU"
complement(18683. .18974)
/rpt_family="ALU"
complement(20349. .20389)
/rpt_family="ALU"
complement(20398. .20501)

repeat_region
/rpt_family="ALU"
complement(21120. .21414)
/rpt_family="ALU"
complement(21722. .25632)
/rpt_family="L1"
23274. .23301
/rpt_family="L1"
26198. .26636
/rpt_family="L1"
complement(27632. .27653)
/rpt_family="L1"
complement(27776. .27803)
/rpt_family="L1"
28127. .28151
/rpt_family="L1"
29781. .29826
/rpt_family="L1"
complement(30309. .30326)
/rpt_family="L1"
complement(30344. .31472)
/rpt_family="L1"
complement(31542. .31829)
/rpt_family="ALU"
complement(32096. .32133)
/rpt_family="L1"
complement(32165. .32729)
/rpt_family="L1"
complement(32758. .33562)
/rpt_family="L1"
complement(34455. .34486)
/rpt_family="L1"
34492. .34778
/rpt_family="ALU"
complement(35297. .35386)
/rpt_family="L1"
35973. .36329
/rpt_family="L1"
complement(36687. .36981)
/rpt_family="ALU"
37162. .37200
/rpt_family="L1"
complement(38651. .38925)
/rpt_family="ALU"
complement(39644. .39687)
/rpt_family="L1"
complement(39688. .39980)
/rpt_family="ALU"
complement(40197. .40226)
/rpt_family="L1"
41164. .41184
/rpt_family="L1"
41716. .41765
/rpt_family="L1"
43720. .43753
/rpt_family="L1"
45223. .45251
/rpt_family="L1"
complement(46066. .46102)
/rpt_family="L1"
complement(46105. .46388)
/rpt_family="ALU"
complement(46748. .47039)
/rpt_family="ALU"
47669. .47958
/rpt_family="ALU"
48675. .49195
/rpt_family="L1"
49988. .50277
/rpt_family="ALU"
complement(50408. .50442)
/rpt_family="L1"
complement(53772. .53879)
/rpt_family="L1"

Single stranded regions:

1

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

| BgIII | | EcoRI | | NsiI | |
|----------|----------|----------|----------|----------|----------|
| Map | Seq | Map | Seq | Map | Seq |
| 1221.00 | 1225.00 | 777.00 | 753.00 | 2434.00 | 2433.00 |
| 9301.00 | 9214.00 | 2717.00 | 2724.00 | 2016.00 | 2016.00 |
| 2260.00 | 2247.00 | 6542.00 | 6545.00 | 958.00 | 925.00 |
| 542.00 | 540.00 | 2717.00 | 2671.00 | 15887.00 | 15655.00 |
| 2421.00 | 2383.00 | 3666.00 | 3619.00 | 1772.00 | 1786.00 |
| 4040.00 | 4054.00 | 2020.00 | 1958.00 | 2230.00 | 2233.00 |
| 1221.00 | 1208.00 | 3319.00 | 3354.00 | 5045.00 | 4995.00 |
| 1068.00 | 1047.00 | 7389.00 | 7370.00 | 527.00 | 518.00 |
| 1482.00 | 1487.00 | 847.00 | 842.00 | 1451.00 | 1421.00 |
| 9301.00 | 9202.00 | 3319.00 | 3252.00 | 958.00 | 945.00 |
| 5493.00 | 5469.00 | 10078.00 | 9931.00 | 5371.00 | 5259.00 |
| 4200.00 | 4199.00 | 5656.00 | 5592.00 | 1002.00 | 994.00 |
| 2816.00 | 2790.00 | 1896.00 | 1892.00 | 5045.00 | 4907.00 |
| 3863.00 | 3891.00 | 10078.00 | 10220.00 | 1072.00 | 1089.00 |
| 2616.00 | 2589.00 | 777.00 | 799.00 | 1002.00 | 999.00 |
| 7106.00 | 7198.00 | 6772.00 | 6793.00 | 3998.00 | 3967.00 |
| 8828.00 | 8842.00 | 2020.00 | 2027.00 | 1532.00 | 1507.00 |
| 5679.00 | 5807.00 | 2311.00 | 2325.00 | 2434.00 | 2427.00 |
| 1389.00 | 1385.00 | 11025.00 | 11100.00 | 1072.00 | 1080.00 |
| 5679.00 | 5685.00 | 8690.00 | 8656.00 | 2016.00 | 2029.00 |
| 13683.00 | 13752.00 | 2897.00 | 2881.00 | 3845.00 | 3825.00 |
| 7777.00 | 7675.00 | 736.00 | 714.00 | 3180.00 | 3192.00 |
| 519.00 | 508.00 | 14213.00 | 14297.00 | 1451.00 | 1448.00 |
| 769.00 | 759.00 | 1041.00 | 1082.00 | 5371.00 | 5459.00 |
| 33833.00 | 30657.00 | 5564.00 | 5453.00 | 646.00 | 632.00 |
| 5276.00 | 5201.00 | 1659.00 | 1648.00 | 748.00 | 737.00 |
| 2816.00 | 2771.00 | 26979.00 | 24815.00 | 4445.00 | 4419.00 |
| 7777.00 | 7750.00 | 6242.00 | 6209.00 | 2326.00 | 2308.00 |
| 2935.00 | 2913.00 | 1041.00 | 1024.00 | 5045.00 | 5023.00 |
| 2816.00 | 2827.00 | 3666.00 | 3630.00 | 8708.00 | 8619.00 |

| | | | | | |
|----------|---------|---------|---------|----------|----------|
| 1221.00 | 1210.00 | 3949.00 | 3923.00 | 527.00 | 531.00 |
| 890.00 | 878.00 | 777.00 | 767.00 | 791.00 | 791.00 |
| 3369.00 | 3317.00 | 1483.00 | 1470.00 | 646.00 | 643.00 |
| 658.00 | 641.00 | 3949.00 | 3851.00 | 3650.00 | 3525.00 |
| 19257.00 | | 5656.00 | 5567.00 | 1532.00 | 1538.00 |
| 2071.00 | | 930.00 | 917.00 | 4804.00 | 4766.00 |
| 2816.00 | | 930.00 | 918.00 | 3650.00 | 3699.00 |
| | | 4033.00 | 3969.00 | 5794.00 | 5678.00 |
| | | 1041.00 | 1017.00 | 17463.00 | 17311.00 |
| | | 8942.00 | | 2690.00 | 2655.00 |
| | | | | 1864.00 | 1835.00 |
| | | | | 1532.00 | 1532.00 |
| | | | | 1451.00 | 1437.00 |
| | | | | 14855.00 | 14780.00 |
| | | | | 6080.00 | 6005.00 |
| | | | | 1532.00 | 1482.00 |
| | | | | 28565.00 | |

Unmatched fragments are due to vector-containing sequences. We expect 3, 1 and 1-2 such fragments in the BgIII, EcoRI, and NsiI enzyme domains, respectively. The sum of these fragments can be calculated from the sequence of the BAC clone and, in this clone, should be 26,152bp, 9,439bp, and 25,563bp for the three enzyme domains.

FEATURES

source

1. .179757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-15"
/clone="De Jong: dj182J23"
/cell_line="Male blood"
/clone_lib="De Jong BAC Library"
2009. .2065
/rpt_family="Alu"
3184. .3272
/rpt_family="MER20"
3891. .4165
/rpt_family="Alu"
complement(4937. .5211)
/rpt_family="Alu"
complement(6618. .6908)
/rpt_family="Alu"
complement(7575. .7871)
/rpt_family="Alu"
complement(8397. .8669)
/rpt_family="Alu"
10475. .10554
/rpt_family="MT1"
complement(13453. .13492)
/rpt_family="Alu"
14584. .14861
/rpt_family="Alu"
complement(15593. .15880)
/rpt_family="Alu"
16455. .16794
/rpt_family="THE1"

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1596: contig of 1596 bp in length
* 1597 2506: gap of unknown length
* 2507 4019: contig of 1513 bp in length
* 4020 4929: gap of unknown length
* 4930 6926: contig of 1997 bp in length
* 6927 7836: gap of unknown length
* 7837 9039: contig of 1203 bp in length
* 9040 9949: gap of unknown length
* 9950 13819: contig of 3870 bp in length
* 13820 14729: gap of unknown length
* 14730 18450: contig of 3721 bp in length
* 18451 19360: gap of unknown length
* 19361 22742: contig of 3382 bp in length
* 22743 23651: gap of unknown length
* 23652 118804: contig of 95153 bp in length
* 118805 119713: gap of unknown length
* 119714 220000: contig of 100287 bp in length.

FEATURES
source 1. 220000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"

BASE COUNT 57281 a 49383 c 48355 g 57702 t 7279 others
ORIGIN

Query Match 16.1%; Score 235; DB 35; Length 220000;
Best Local Similarity 82.2%; Pred. No. 5.8e-45;
Matches 282; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

OY 1118 TGAACACGAGAGAAACAGGAAATCAAGACTCTACAAAGAGCGTGGCAGGTGGCT 1177
Db 165051 TGCACCTTGAAACATTTCTGCCAAGTGAAGAACCACAGACAGCGCGGCATGTGGCT 165110
OY 1178 CATGCCTGTAATCCAGCACTTTGGAGG-CGAGGCAGGCAGATCACTTGAGSTAAGAG 1236
Db 165111 CATGCCTGTAATCCAGCACTGGGGAGGCGCGGGTGGATCACTGAGGTTGGGAG 165170
OY 1237 TTCAAGACCAGCCTGGCCAAATGGTAATCCTGTCTACTATAAATACAAAAGTTAG 1296
Db 165171 TTCAAGACCAGCCTGGCCAAACATGGTGAACCCCTGTCTCTACTATAAATAGAAAAATAG 165230
OY 1297 CTGATATGTTGGCAGGCGCCTGTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATT 1356
Db 165231 CCGGGTGTGCTAGAGGACACCTGTAATCCAGCTACTCGGAGGCTGAGGCAGAGAAATT 165290
OY 1357 GCTGAATATGGAGGCGAGAGGTTGAAGTGAAGTGAATCAGACCACTATCTCCAGCTG 1416
Db 165291 GCTTGAACCCAGGAGGAGAGAGTTGCAAGTGGGCTGAGATCTCGCAGCTGCACCTCCAGCT 165350
OY 1417 GGGCAACAGAGTAAGACTCTGTCTCAAAAAAATAAAAAA 1459
Db 165351 GGTCAACAGAGTGAACCTGTCTCAAAAAAATAAAAAA 165393

RESULT 13
HS30M3/c 102200 bp DNA PRI 28-JAN-1999
LOCUS Human DNA sequence from clone 30M3 on chromosome 6p22.1-22.3.
DEFINITION Contains three novel genes, one similar to C. elegans Y63D3A.4 and
one similar to (predicted) plant, worm, yeast and archaea bacterial
genes, and the first exon of the KIAA0319 gene. Contains ESTs, GSS
and putative CpG islands, complete sequence.
ACCESSION AL031775
NID 94071041
VERSION AL031775.1 GI:4071041
KEYWORDS HTG; CpG island; KIAA0319; Y63D3A.4.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 102200)
TITLE Phillips,S.
JOURNAL Direct Submission

COMMENT

Submitted (05-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Dec 29, 1998 this sequence version replaced gi:4007165.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence is the entire insert of clone 30M3.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
30M3 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://dpcpac.med.buffalo.edu/VECTOR:pcypac2.

FEATURES

source 1. 102200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="30M3"
/map="p22.1-22.3"
/clone_1lb="RPC11"
757..825
/note="MER47 repeat: matches 9. .76 of consensus"
829..932
/note="MER47 repeat: matches 2226. .2319 of consensus"
905..976
/note="MER47 repeat: matches 2249. .2323 of consensus"
1320..1641
/note="AluSc repeat: matches 1. .299 of consensus"
1642..1927
/note="AluDb repeat: matches 1. .289 of consensus"
2392..2517
/note="FLAM_C repeat: matches 1. .126 of consensus"
2760..2789
/note="15 copies 2 mer tg 97% conserved"
3013..3313
/note="AluDo repeat: matches 1. .301 of consensus"
3467..3763
/note="AluSp repeat: matches 1. .299 of consensus"
complement(3650..4203)
/note="match: GSS B63434"
4458..4591
/note="MIR repeat: matches 53. .208 of consensus"
4799..4952
/note="MER3 repeat: matches 3. .154 of consensus"
join(<5119..5253,7768..9169)
/gene="dJ30M3.2"
/note="match: ESTs AA259531 AA289775 NS0863 AA806198
AA931087 AA830274 AI199746 AA912273 AI081164 AI288433
AA604069 AA772465 AA876636 AA522552 AI147796 AA639167
AA655259 AA289775 AA259531 AA731057 AA290469 D20450
AA517053 AA669103 AA974142 AA903560 AI083595 AI287615
AI041875 AI243439 NA5220 R25494 AA798467 AA798123 AA591228
AA473082 AI120905 AA863774 AA840219 AA041897 D77719
AI019712 AI006885 AI194277 D56188 AA926179 AA851214 T51937


```

gene
CDS
    R01243; supported by GENSCAN"
    /evidence=not_experimental
    /product="dJ30M3.2 (novel protein)"
    5119. .9169
    /gene="dJ30M3.2"
    join(<5119. .5253,7768. .7893)
    /gene="dJ30M3.2"
    /note="supported by GENSCAN"
    /codon_start=1
    /evidence=not_experimental
    /product="dJ30M3.2 (novel protein)"
    /protein_id="CAA21140.1"
    /db_xref="PID:e1373176"
    /db_xref="PID:g4210352"
    /db_xref="GI:4210352"
    /translation="AKFEFHGDYEQFLHVLRSKDTGIVNNPNQSVFLFDROHL
    QTPKRNATIFKLSICLYLPQQLTHWAVGTIEDHLPYME"
    5393. .5674
    /note="AluSg repeat: matches 1. .291 of consensus"
    5715. .5763
    /note="MER5B repeat: matches 1. .49 of consensus"
    5883. .6000
    /note="MER5A repeat: matches 37. .189 of consensus"
    6010. .6308
    /note="AluY repeat: matches 1. .295 of consensus"
    6384. .6667
    /note="AluJo repeat: matches 5. .296 of consensus"
    6740. .6878
    /note="MIR repeat: matches 69. .207 of consensus"
    7106. .7400
    /note="AluJb repeat: matches 1. .297 of consensus"
    8676. .8721
    /note="23 copies 2 mer tt 78% conserved"
    9092. .9129
    /note="19 copies 2 mer aa 82% conserved"
    9124. .9129
    /gene="dJ30M3.2"
    9169
    /gene="dJ30M3.2"
    9439. .9621
    /note="MIR repeat: matches 8. .256 of consensus"
    9691. .9798
    /note="UIMC4 repeat: matches 7851. .7977 of consensus"
    10020. .10149
    /note="UIMD3 repeat: matches 7361. .7499 of consensus"
    10507. .10716
    /note="UIMC4 repeat: matches 6633. .6858 of consensus"
    10855. .11115
    /note="AluJo repeat: matches 1. .283 of consensus"
    11469. .11619
    /note="UIM1A1 repeat: matches 1. .146 of consensus"
    11620. .11906
    /note="AluSx repeat: matches 1. .287 of consensus"
    11907. .12055
    /note="UIM1A1 repeat: matches 146. .310 of consensus"
    complement(join(12316. .12571,15963. .>16147))
    /gene="dJ30M3.1"
    /note="mach: ESTS AA522110 AA207689 AA819653 AA928745
    AI017897 AA993303 AI216348 AA832106 AA121362 AA552456
    AA629986 AA614605 T55234 AA121362 F17737 AA445103 AA073807
    AA121372 AI052396 AA939005 AI267851 T53974; supported by
    GENSCAN and FGENES"
    /evidence=not_experimental
    /product="dJ30M3.1 (novel protein similar to (predicted)
    plant, worm, yeast and archaea bacterial proteins)"
    complement(12316)
    /gene="dJ30M3.1"
    complement(12316. .16147)
    /gene="dJ30M3.1"
    complement(12331. .12336)
    /gene="dJ30M3.1"
    complement(join(12415. .12571,15963. .>16147))
    /gene="dJ30M3.1"

```

[illegible]

QY 1439 CTCAAAAAAAAAAAAAAAAAAAA 1459
|||||
Db 83524 CTCAAAAAAAAAAAAAAAAAAAA 83504

RESULT 14
AC004024 80507 bp DNA PRI 17-SEP-1998
LOCUS Homo sapiens 12q24.2 PAC RPi1-128M12 (Roswell Park Cancer
DEFINITION Institute Human PAC library) complete sequence.
AC004024
AC004024
AC004024.1 GI:3366553
93366553
VERSION HTG.
KEYWORDS AC004024.1 GI:3366553
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 80507)
AUTHORS Muzny,D., Aronson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M.,
Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpachy,S., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
Lu,J., Ly,T., Marondel,I., Martinez,C., Merscher,S., Montgomery,K.,
Oswal,G., Perez,L., Rashid,N.D., Renault,B., Rowland,K., Savage,L.,
Scherer,S.E., Shen,H., Simon,M., Stovall,K., Timms,K.M., Todd,J.,
Vo,O., Williamson,A., Worley,K.C., Yu,W., Kucherlapati,R.,
Nelson,D. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 80507)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 80507)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 80507)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 80507)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 80507)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 31, 1998 this sequence version replaced gi:3242678.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
FEATURES
source location/Qualifiers
1. 80507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC11-120M12"
/chromosome="12q24.2"
1. 107
/note="Overlaps with bases 139060 to 139166 in AC003029."
/note="Region: Overlap with AC003029."
239. 540
/rpt_family="AluSg"
541. 663
/rpt_family="(GGA)n"
666. 749
/rpt_family="(GAA)n"
763. 1059
/rpt_family="AluSx"
complement(join(2358. 2887,3646. 3719,4946. 5090,
7752. 7890,8220. 8254,38145. 38168,64167. 64298))
/gene="Unknown"
/note="Contains two Unigene clusters. Representative
accession numbers are AA437268, N29463, and clone
za72d06.r1."
2391. 2559
/gene="Unknown"
/standard_name="G19891"
/note="Genethon-pending:RH10692, A002B25, STS32769,
GDB:4563466, Chr. 12-12, Homo Sapiens"
/db_xref="dbSTS:32769"
4019. 4317
/rpt_family="AluSx"
complement(4547. 4855)
/rpt_family="AluY"
5542. 5836
/rpt_family="AluSg"
complement(6157. 6456)
/rpt_family="AluSx"
complement(6674. 6720)
/rpt_family="(CA)n"
complement(6721. 6998)
/rpt_family="AluSx"
complement(7363. 7392)
/rpt_family="AT-rich"
10233. 10531
/rpt_family="AluSc"
10576. 10907
/rpt_family="AluJo"
complement(10941. 10998)
/rpt_family="(CA)n"
11051. 11338
/rpt_family="AluSx"
complement(11339. 11396)
/rpt_family="(TA)n"
complement(11421. 11516)
/rpt_family="AluJo/FRAM"
complement(11526. 11644)
/rpt_family="(TA)n"
12156. 12452
/rpt_family="AluSc"
complement(12512. 12693)
/rpt_family="MER3"
complement(12998. 13296)
/rpt_family="AluSp"
complement(13329. 13631)
/rpt_family="AluY"
complement(13639. 13726)
/rpt_family="MIR"
complement(13819. 13874)
/rpt_family="(CA)n"
complement(14076. 14283)


```
* 37005 40411: contig of 3407 bp in length
* 40412 40430: gap of unknown length
* 40431 50114: contig of 9684 bp in length
* 50115 50133: gap of unknown length
* 50134 67886: contig of 17753 bp in length
* 67887 67905: gap of unknown length
* 67906 84164: contig of 16259 bp in length
* 84165 84183: gap of unknown length
* 84184 105583: contig of 21400 bp in length
* 105584 105602: gap of unknown length
* 128037 128036: contig of 22434 bp in length
* 128056 128055: gap of unknown length
* 128056 185122: contig of 57067 bp in length
* 185123 185140: gap of unknown length
* 185141 186840: contig of 1700 bp in length
* 186841 186858: gap of unknown length
* 186859 188477: contig of 1619 bp in length.
```

FEATURES
source
1. .188477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1125K23"

BASE COUNT 43810 a 48919 c 48642 g 46728 t 378 others
ORIGIN

Query Match 16.1%; Score 234.6; DB 34; Length 188477;
Best Local Similarity 84.3%; Pred. No. 7.1e-45;
Matches 264; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
OY 1146 AAGACTCTACAAAGAGGCTGGGCAGGGTGGCTCATGCTGTAATCCAGCACTTTGGGAG 1205
Db 86621 AGGATATTGCAAAAGGGCTGGGCACGGTGGCTCACAACCTGTAATCCAGCACTTTGGAGG 86680

OY 1206 GCGAGGCGAGCAGATCACTTGAGTTAAGAGTTCAAGACCAGCCTGGCCAAAATGGTGAA 1265
Db 86681 CCAAGCGGGCAGATCACTTGAGTTGGAGTTCAAGACCAGCCTGGGCCAATGATGTGAA 86740

OY 1266 ATCCTGTCTGTACTAAATAACAAAGTTAGCTGATATGTGGCAGGCGCTGTAATCC 1325
Db 86741 ACCCTGTCTCTACTAAATAACAAATATATCGACGTGTGGCGGGCACCCTGTAATCC 86800

OY 1326 CAGCTACTTGGGAGGCTGAGGCGAGGAGATTCCTGAATATGGGAGGAGAGGTTGAAGT 1385
Db 86801 CAGCTACTTGGGAGGCTGAGGCGAGGAGATTCCTGAAGCTTGGGAGGAGGTTGTGGT 86860

OY 1386 GAGTTGAGATCACACCACTATCTCCAGCTGGGGCAACAGAGTAAGACTCTGTCTCAAAA 1445
Db 86861 GAGCCGAGATGCCGCCACTGCACCTCCAGCCTGGGCCAAGAGAGTGAGACCTGTCTCAAAA 86920

OY 1446 AAAAAAAAAAAAAA 1458
Db 86921 AAAAAAAAAAAAAA 86933
```

Search completed: September 25, 1999, 23:36:02
Job time: 7305 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:56:30 ; Search time 273 Seconds
(without alignments)
1337.107 Million cell updates/sec

Title: US-09-030-606-174

Perfect score: 1459
Sequence: 1 GGTCAGCCGACACTGTTC.....TCAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 1456 | 99.8 | 1459 | 1 | V58646 | Prostate tumour sp |
| 2 | 1456 | 99.8 | 1459 | 1 | V61251 | CDNA sequence of p |
| 3 | 474 | 32.5 | 1265 | 1 | V58645 | Prostate tumour sp |
| 4 | 474 | 32.5 | 1265 | 1 | V61250 | CDNA sequence of p |
| 5 | 329 | 22.5 | 1386 | 1 | V11855 | Homo sapiens Tub I |
| 6 | 329 | 22.5 | 1248 | 1 | V58644 | Prostate tumour sp |
| 7 | 329 | 22.5 | 1248 | 1 | V61249 | CDNA sequence of p |
| 8 | 292.8 | 20.1 | 1167 | 1 | V58647 | Prostate tumour sp |
| 9 | 292.8 | 20.1 | 1167 | 1 | V61252 | CDNA sequence of p |
| 10 | 291.8 | 20.0 | 871 | 1 | V37495 | Human prostate-spe |
| 11 | 289.4 | 19.8 | 402 | 1 | X41114 | Human secreted pro |
| 12 | 257.4 | 17.6 | 1119 | 1 | V58648 | Prostate tumour sp |
| 13 | 257.4 | 17.6 | 1119 | 1 | V61253 | CDNA sequence of p |
| 14 | 226.2 | 15.5 | 24026 | 1 | T18325 | BRCA1, human breas |
| 15 | 226.2 | 15.5 | 24025 | 1 | T17455 | Mutated BRCA1 geno |
| 16 | 226.2 | 15.5 | 24026 | 1 | T17512 | Mutated BRCA1 geno |
| 17 | 226.2 | 15.5 | 24026 | 1 | T17513 | Mutated BRCA1 geno |
| 18 | 226.2 | 15.5 | 24026 | 1 | T17514 | Mutated BRCA1 geno |
| 19 | 226.2 | 15.5 | 24025 | 1 | T17515 | Mutated BRCA1 geno |
| 20 | 226.2 | 15.5 | 24026 | 1 | T17516 | Mutated BRCA1 geno |
| 21 | 226.2 | 15.5 | 24026 | 1 | T17517 | Mutated BRCA1 geno |
| 22 | 226.2 | 15.5 | 24026 | 1 | T17518 | Mutated BRCA1 geno |
| 23 | 226.2 | 15.5 | 24026 | 1 | T17519 | Mutated BRCA1 geno |
| 24 | 226.2 | 15.5 | 24029 | 1 | T17520 | Mutated BRCA1 geno |
| 25 | 226.2 | 15.5 | 24026 | 1 | T17521 | Mutated BRCA1 geno |
| 26 | 226.2 | 15.5 | 24026 | 1 | T17522 | Mutated BRCA1 geno |
| 27 | 226.2 | 15.5 | 24026 | 1 | T17523 | Mutated BRCA1 geno |
| 28 | 226.2 | 15.5 | 24026 | 1 | T17524 | Mutated BRCA1 geno |
| 29 | 226.2 | 15.5 | 24031 | 1 | T17525 | Mutated BRCA1 geno |
| 30 | 226.2 | 15.5 | 24026 | 1 | T17526 | Mutated BRCA1 geno |
| 31 | 226.2 | 15.5 | 24026 | 1 | T17527 | Mutated BRCA1 geno |
| 32 | 226.2 | 15.5 | 24026 | 1 | T17528 | Mutated BRCA1 geno |
| 33 | 226.2 | 15.5 | 24026 | 1 | T17529 | Mutated BRCA1 geno |
| 34 | 226.2 | 15.5 | 24026 | 1 | T17530 | BRCA1 genomic sequ |
| 35 | 226.2 | 15.5 | 24026 | 1 | T32612 | BRCA1, human breas |
| 36 | 223.2 | 15.3 | 50000 | 1 | X23517 | Human kidney amino |
| 37 | 219.2 | 15.0 | 7676 | 1 | V62906 | Human galactokinas |
| 38 | 218.6 | 15.0 | 7849 | 1 | Q94109 | hML genomic DNA. T |
| 39 | 216.8 | 14.9 | 452 | 1 | T42809 | Polymorphic locus |
| 40 | 215.4 | 14.8 | 1618 | 1 | Q10207 | PH13C insert conta |
| 41 | 215.4 | 14.8 | 1618 | 1 | Q46958 | Human cytokine syn |
| 42 | 215.4 | 14.8 | 3234 | 1 | Q92781 | Human thymopoietin |
| 43 | 213.4 | 14.6 | 20303 | 1 | T71699 | Human deoxycytidy |

ALIGNMENTS

44 213.4 14.6 26764 1 T71696
45 213.4 14.6 15056 1 V52967

Human deoxycytidy
Carinoembryonic a

RESULT 1
ID V58646
AC V58646; standard; CDNA; 1459 BP.
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
be used in the method of the invention. The method is for detecting
prostate cancer comprises contacting a biological sample with an agent
able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 99.8%; Score 1456; DB 1; Length 1459;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTCAGCCGACACTGTTCCAGAGTGAGTGACAGCTCTACACCATCGGCTGGCC 60
DB 1 GGTCAGCCGACACTGTTCCAGAGTGAGTGACAGCTCTACACCATCGGCTGGCC 60
QY 61 TGACAGTCTTGAGCGCCAGACCAAGAGCCAGGAGCCAGATGTGAGGCGCCTCTCCG 120
DB 61 TGACAGTCTTGAGCGCCAGACCAAGAGCCAGGAGCCAGATGTGAGGCGCCTCTCCG 120
QY 121 TACGGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATGCTCAAGTTGG 180
DB 121 TACGGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATGCTCAAGTTGG 180
QY 181 ACGAATCCGTGTCGAGTTCGACACCATCCGGAGCATGACATGCTTGCGAGTGGCCCTTA 240
DB 181 ACGAATCCGTGTCGAGTTCGACACCATCCGGAGCATGACATGCTTGCGAGTGGCCCTTA 240
QY 241 CCGCGGGGAACCTCTTGCCCTGTTTCTGGCTGGGGTCTGCGGAACGGTGAGCTCACGG 300
DB 241 CCGCGGGGAACCTCTTGCCCTGTTTCTGGCTGGGGTCTGCGGAACGGTGAGCTCACGG 300
QY 301 GTGTGTCTGCCCCCTTCAAGAGGTCTCTGCCAGTCGCGGGGGCTGACCCAGAGCT 360
DB 301 GTGTGTCTGCCCCCTTCAAGAGGTCTCTGCCAGTCGCGGGGGCTGACCCAGAGCT 360
QY 361 CTGCTCCAGGAGAGATGCTACCGTGTGACATGCGCTGACAGCTGTGCTGTCTGA 420
DB 361 CTGCTCCAGGAGAGATGCTACCGTGTGACATGCGCTGACAGCTGTGCTGTCTGA 420
QY 421 NGAGGTCTGCANTAGCTTATGACCCCGCTGTACCACCCAACATGTTTGGCGCGCGG 480

|||||
Db 421 NGAGGTCTGCANTAACTCTATGACCCCGCTGTACCACCCCANCATGTTCTGCGCGCGG 480
QY 481 AGGGCAAGACCAGAAGACTCTCTGCACAGTGAGAGAGGGGAAAGGGGCGAGCGGACT 540
|||||
Db 481 AGGGCAAGACCAGAAGACTCTCTGCACAGTGAGAGAGGGGAAAGGGGCGAGCGACT 540
QY 541 CAGGGAAGGGTGAGAGAGGGGAGACAGAGACACAGGGCCGCATGGCGAGATGCAGAG 600
|||||
Db 541 CAGGGAAGGGTGAGAGAGGGGAGACAGAGACACAGGGCCGCATGGCGAGATGCAGAG 600
QY 601 ATGGAGAGACACACAGGAGACAGTGCACACTAGAGAGAACTGAGAGAAACAGAGAA 660
|||||
Db 601 ATGGAGAGACACACAGGAGACAGTGCACACTAGAGAGAACTGAGAGAAACAGAGAA 660
QY 661 ATAAACACAGGAATAAAGAGAACCAAGAGAGAGAAACAGAAACAGACATGGGAGGC 720
|||||
Db 661 ATAAACACAGGAATAAAGAGAACCAAGAGAGAGAAACAGAAACAGACATGGGAGGC 720
QY 721 AGAAACACACACACATAGAAATGCAGTTGACCTTCCACACAGCATGGGCGCTGAGGCGGT 780
|||||
Db 721 AGAAACACACACACATAGAAATGCAGTTGACCTTCCACACAGCATGGGCGCTGAGGCGGT 780
QY 781 GACCTCCACCCCAATAGAAATCCTCTTATTAACCTTTGACTCCCCAAAAACCTGACTAGAA 840
|||||
Db 781 GACCTCCACCCCAATAGAAATCCTCTTATTAACCTTTGACTCCCCAAAAACCTGACTAGAA 840
QY 841 ATAGCCTACTGTGACGGGGAGCCTTACCATAACATAATAGTCGATTATGCATACGT 900
|||||
Db 841 ATAGCCTACTGTGACGGGGAGCCTTACCATAACATAATAGTCGATTATGCATACGT 900
QY 901 TTTATGCATTCATGATATACCTTTGTTGGAATTTTGTATTTCTAAGCTACACAGTTC 960
|||||
Db 901 TTTATGCATTCATGATATACCTTTGTTGGAATTTTGTATTTCTAAGCTACACAGTTC 960
QY 961 GTCTGTGAATTTTAAATTTGTTGCAACTCTCTTAATAATTTTCTGATGTTATTGA 1020
|||||
Db 961 GTCTGTGAATTTTAAATTTGTTGCAACTCTCTTAATAATTTTCTGATGTTATTGA 1020
QY 1021 AAAAAATCCAGTATAGTGAAGTGTGCATTCAAACAGGGTGTTCAGGGTCAACTGT 1080
|||||
Db 1021 AAAAAATCCAGTATAGTGAAGTGTGCATTCAAACAGGGTGTTCAGGGTCAACTGT 1080
QY 1081 GTACCCAGAGGGAACAGTACACAGATTCATAGAGTGAACACGAAAGAGAAACAGGAA 1140
|||||
Db 1081 GTACCCAGAGGGAACAGTACACAGATTCATAGAGTGAACACGAAAGAGAAACAGGAA 1140
QY 1141 AAATCAAGACTCTACAAGAGCCTGGGAGGGTGGCTCATGCTGTATCCAGCACTTT 1200
|||||
Db 1141 AAATCAAGACTCTACAAGAGCCTGGGAGGGTGGCTCATGCTGTATCCAGCACTTT 1200
QY 1201 GGGAGGCGAGGAGGAGAGTCACTTGAGGTAAGGATTCAGAGCCAGCCTGGCCAAATG 1260
|||||
Db 1201 GGGAGGCGAGGAGGAGAGTCACTTGAGGTAAGGATTCAGAGCCAGCCTGGCCAAATG 1260
QY 1261 GTGAAATCCTGTCTGTACTTAAATAACAAAAGTTAGCTGATATGTTGCGAGCGCCTGT 1320
|||||
Db 1261 GTGAAATCCTGTCTGTACTTAAATAACAAAAGTTAGCTGATATGTTGCGAGCGCCTGT 1320
QY 1321 AATCCAGCTACTTGGAGGCTGAGGAGGAGAAATGCTTGAATATGGAGGCGAGAGTT 1380
|||||
Db 1321 AATCCAGCTACTTGGAGGCTGAGGAGGAGAAATGCTTGAATATGGAGGCGAGAGTT 1380
QY 1381 GAAGTGAGTTGAGATCACACCACTATACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCT 1440
|||||
Db 1381 GAAGTGAGTTGAGATCACACCACTATACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCT 1440
QY 1441 CAAAAAAG 1459
|||||
Db 1441 CAAAAAAG 1459

RESULT - 2*

V61251
ID V61251 standard; cDNA; 1459 BP.
AC V61251.
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 99.8%; Score 1456; DB 1; Length 1459;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGAGCCGACACACTGTTTCCAGAAGTGAAGTGCAGAGCTCTACACCATCGGGCTGGGCC 60
|||||
Db 1 GGTGAGCCGACACACTGTTTCCAGAAGTGAAGTGCAGAGCTCTACACCATCGGGCTGGGCC 60
QY 61 TGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCG 120
|||||
Db 61 TGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCG 120
QY 121 TACGGCACCCAGAGTACACAGAGCCCTTGTCTGCTTAACGACCTCATGCTCAAGTTGG 180
|||||
Db 121 TACGGCACCCAGAGTACACAGAGCCCTTGTCTGCTTAACGACCTCATGCTCAAGTTGG 180
QY 181 ACGAATCCGTGTCGAGTGTACACCATCCGAGCATCAGCATGCTTGCAGTGCCCTA 240
|||||
Db 181 ACGAATCCGTGTCGAGTGTACACCATCCGAGCATCAGCATGCTTGCAGTGCCCTA 240
QY 241 CCGCGGGAACTCTTGCTCTTCAAGAGGTCTCTGCCCCAGTCGGGGGCTGACCCAGAGCT 300
|||||
Db 241 CCGCGGGAACTCTTGCTCTTCAAGAGGTCTCTGCCCCAGTCGGGGGCTGACCCAGAGCT 300
QY 301 GTGTGTCTTGCCCTCTTCAAGAGGTCTCTGCCCCAGTCGGGGGCTGACCCAGAGCT 360
|||||
Db 301 GTGTGTCTTGCCCTCTTCAAGAGGTCTCTGCCCCAGTCGGGGGCTGACCCAGAGCT 360
QY 361 CTGCGTCCAGGAGAGATGCTTACCCTGCTGCTGAGTGGTGAACGTTGCTGTTGA 420
|||||
Db 361 CTGCGTCCAGGAGAGATGCTTACCCTGCTGCTGAGTGGTGAACGTTGCTGTTGA 420
QY 421 NGAGGTCTGCANTAACTCTATGACCCGCTGTACCAACCCCANCATGTTCTGCGCGCGG 480
|||||
Db 421 NGAGGTCTGCANTAACTCTATGACCCGCTGTACCAACCCCANCATGTTCTGCGCGCGG 480
QY 481 AGGGCAAGACCAGAAGACTCTCTGCAACGTTGAGAGAGGGGAAAGGGGAGGCGAGGACT 540
|||||
Db 481 AGGGCAAGACCAGAAGACTCTCTGCAACGTTGAGAGAGGGGAAAGGGGAGGCGAGGACT 540
QY 541 CAGGGAAGGGTGAGAGAGGGGAGACAGAGACACAGGGCCGATGGCGAGATGCAGAG 600
|||||
Db 541 CAGGGAAGGGTGAGAGAGGGGAGACAGAGACACAGGGCCGATGGCGAGATGCAGAG 600
QY 601 ATGGAGAGACACACAGGAGAGACAGTGCACACTAGAGAGAGAACTGAGAGAAACAGAGAA 660

| | | | |
|----|------|---|------|
| Db | 601 | ATGAGAGACACACAGGGAGACAGTGACAACTAGAGAGAGAACTGAGAGAAACAGAGAA | 660 |
| QY | 661 | ATAAACACAGGAATTAAGAGAGCAAAGAGAGAGAAACAGAAACAGACATGGGAGCC | 720 |
| Db | 661 | ATAAACACAGGAATTAAGAGAGCAAAGAGAGAGAGAAACAGAAACAGACATGGGAGCC | 720 |
| QY | 721 | AGAAACACACACACATAGAAATGCAGTTGACCTTCCAAACAGCATGGGGCCTGAGGGCGGT | 780 |
| Db | 721 | AGAAACACACACACATAGAAATGCAGTTGACCTTCCAAACAGCATGGGGCCTGAGGGCGGT | 780 |
| QY | 781 | GACCTCCACCCCAATAGAAAATCCTCTTATTAACCTTTTGACTCCCAAAAAACCTGACTAGAA | 840 |
| Db | 781 | GACCTCCACCCCAATAGAAAATCCTCTTATTAACCTTTTGACTCCCAAAAAACCTGACTAGAA | 840 |
| QY | 841 | ATAGCCTACTGTTGACGGGGAGCCCTTACCAATAACATAATAGTCGATTATGCATACGT | 900 |
| Db | 841 | ATAGCCTACTGTTGACGGGGAGCCCTTACCAATAACATAATAGTCGATTATGCATACGT | 900 |
| QY | 901 | TTTATGCATTCATGATATACCTTTGTTGGAATTTTGTGATATTTCTAAGCTACACAGTTC | 960 |
| Db | 901 | TTTATGCATTCATGATATACCTTTGTTGGAATTTTGTGATATTTCTAAGCTACACAGTTC | 960 |
| QY | 961 | GTCGTGTAATTTTAAATGTTGCAACTCTCCATAATTTTCTGATGTGTTATTGA | 1020 |
| Db | 961 | GTCGTGTAATTTTAAATGTTGCAACTCTCCATAATTTTCTGATGTGTTATTGA | 1020 |
| QY | 1021 | AAAAATCCAGTATATAGTGAGACTTGTGCATTCAAACACAGGGTGTTCAGGGTCACTGT | 1080 |
| Db | 1021 | AAAAATCCAGTATATAGTGAGACTTGTGCATTCAAACACAGGGTGTTCAGGGTCACTGT | 1080 |
| QY | 1081 | GTACCCAGAGGGAACAGTGACACACAGATTTCATAGAGGTGAACACGAAAGAGAAACAGGAA | 1140 |
| Db | 1081 | GTACCCAGAGGGAACAGTGACACACAGATTTCATAGAGGTGAACACGAAAGAGAAACAGGAA | 1140 |
| QY | 1141 | AAATCAAGACTCTACAAAGAGGCTGGGCAGGGTGGCTCATGCCGTATCCACAGCACTTT | 1200 |
| Db | 1141 | AAATCAAGACTCTACAAAGAGGCTGGGCAGGGTGGCTCATGCCGTATCCACAGCACTTT | 1200 |
| QY | 1201 | GGGAGGCGAGGCGAGGCGAGATCACTTGAGGTAAGGATTCAGAACCAGCCCTGGCCAAATG | 1260 |
| Db | 1201 | GGGAGGCGAGGCGAGGCGAGATCACTTGAGGTAAGGATTCAGAACCAGCCCTGGCCAAATG | 1260 |
| QY | 1261 | GTGAATCTGTCTGTACTAAAAATACAAAAAGTTAGCTGGATATGGTGGCAGGGCCTGT | 1320 |
| Db | 1261 | GTGAATCTGTCTGTACTAAAAATACAAAAAGTTAGCTGGATATGGTGGCAGGGCCTGT | 1320 |
| QY | 1321 | AATCCAGCTACTTGGAGGCTGAGGCGAGGAAATTCCTGAATATGGAGGCGAGAGGTT | 1380 |
| Db | 1321 | AATCCAGCTACTTGGAGGCTGAGGCGAGGAAATTCCTGAATATGGAGGCGAGAGGTT | 1380 |
| QY | 1381 | GAAGTGAGTTGAGATCACACCACCTATACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCT | 1440 |
| Db | 1381 | GAAGTGAGTTGAGATCACACCACCTATACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCT | 1440 |
| QY | 1441 | CAAAAAAAAAAAAAAAAAA 1459 | |
| Db | 1441 | CAAAAAAAAAAAAAAAAAA 1459 | |

| RESULT | 3 |
|--------|---|
| V58645 | |
| ID | V58645 standard; cDNA; 1265 BP. |
| AC | V58645; |
| DT | 08-DEC-1998 (first entry) |
| DE | Prostate tumour specific gene clone DE2. |
| KW | Prostate tumour specific gene; human; prostate cancer; detection; |
| KW | therapy; ss. |
| OS | Homo sapiens. |
| PN | W09837418-A2. |
| PD | 27-AUG-1998. |
| PF | 25-FEB-1998; U03690. |
| PR | 09-FEB-1998; U5-904809. |

PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

| | | | | |
|-----------------------|--------------|--------------------|---------------|--------------|
| Query Match | 32.5%; | Score 474; | DB 1; | Length 1265; |
| Best Local Similarity | 98.4%; | Pred. No. 4.4e-95; | | |
| Matches 477; | Conservative | 0; | Mismatches 8; | Indels 0; |
| | | | | Gaps 0 |

| | | | |
|----|-----|--|-----|
| QY | 25 | AGTAGTGCAGAGCTCCTACACCACTCGGGCTGGGGCTGCACAGTCTTGAGGCCGACCAAG | 84 |
| | | | |
| Db | 104 | ACTGTTCCAGAACTCCTACACCACTCGGGCTGGGGCTGCACAGTCTTGAGGCCGACCAAG | 163 |
| QY | 85 | AGCCAGGGAGCCAGATGGTGGAGGGCCAGCCTCTCCCGTACGGCACCAGAGTACACAGAC | 144 |
| | | | |
| Db | 164 | AGCCAGGGAGCCAGATGGTGGAGGGCCAGCCTCTCCCGTACGGCACCAGAGTACACAGAC | 223 |
| QY | 145 | CCTTGCTCGTAAACGACCCTCATGCTCATCAAGTTGAGCAAATCCGTGCCAGTGTGACA | 204 |
| | | | |
| Db | 224 | CCTTGCTCGTAAACGACCCTCATGCTCATCAAGTTGAGCAAATCCGTGCCAGTGTGACA | 283 |
| QY | 205 | CCATCCGGAGCATCAGCATTGCTTCGACAGTGCCCTACCGCGGGGAACCTTGCCCTGTT | 264 |
| | | | |
| Db | 284 | CCATCCGGAGCATCAGCATTGCTTCGACAGTGCCCTACCGCGGGGAACCTTGCCCTGTT | 343 |
| QY | 265 | CTGGCTGGGCTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTTTCAAGA | 324 |
| | | | |
| Db | 344 | CTGGCTGGGCTCTGCTGGCGAAGCGGTGAGCTCACGGGTGTGTCTGCCCTTTCAAGA | 403 |
| QY | 325 | GGTCTCTGCCCCAGTCGCGGGGGGTGACCCAGAGCTCTGCGTCCCAGGAGATGCTTAC | 384 |
| | | | |
| Db | 404 | GGTCTCTGCCCCAGTCGCGGGGGGTGACCCAGAGCTCTGCGTCCCAGGAGATGCTTAC | 463 |
| QY | 385 | CGTGCTGCAGTGCGTGAACGTGTGCGTGGTGTCTGTGANGAGGTCTGCANTAAAGCTTATGA | 444 |
| | | | |
| Db | 464 | CGTGCTGCAGTGCGTGAACGTGTGCGTGGTGTCTGTGANGAGGTCTGCANTAAAGCTTATGA | 523 |
| QY | 445 | CCCGCTGTACACCCCAACATGTCTGCGCGGGCGGAGGGCAAGACAGAGGACTCTG | 504 |
| | | | |
| Db | 524 | CCCGCTGTACACCCCAACATGTCTGCGCGGGCGGAGGGCAAGACAGAGGACTCTG | 583 |
| QY | 505 | CAACG 509 | |
| | | | |
| Db | 584 | CAACG 588 | |

| RESULT | 4 |
|--------|---|
| V61250 | |
| ID | V61250 standard; cDNA; 1265 BP. |
| AC | V61250; |
| DT | 06-JAN-1999 (first entry) |
| DE | CDNA sequence of prostate tumour clone P703 splice variant DE2. |
| KW | Prostate; cancer; tumour; vaccine; immunogen; clone; ss. |
| OS | Homo sapiens. |
| PN | W09837093-A2. |
| PD | 27-AUG-1998. |
| PF | 25-FEB-1998; U03492. |
| PR | 09-FEB-1998; U5-020956. |

PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
CC Claim 3; Page 105-106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.5%; Score 474; DB 1; Length 1265;
Best Local Similarity 98.4%; Pred. No. 4.4e-95;
Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 25 AGTGAGTCAGAGCTCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 84
DB 104 ACTGTTCCAGAACTCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 163
QY 85 AGCCAGGAGCCAGATGTGAGGCGCAGCTCTCCGTACGGCACCACCAAGTACACAGAC 144
DB 164 AGCCAGGAGCCAGATGTGAGGCGCAGCTCTCCGTACGGCACCACCAAGTACACAGAC 223
QY 145 CCTGTGCTGCTAAGCACTCATGCTCATCAAGTTGAGCAATCCGTCCGAGTCTGACA 204
DB 224 CCTGTGCTGCTAAGCACTCATGCTCATCAAGTTGAGCAATCCGTCCGAGTCTGACA 283
QY 205 CCATCCGAGCATCAGATTTGCTTGCAGTGCCTTACCCTGCGGGAATCTTGCTCTGTT 264
DB 284 CCATCCGAGCATCAGATTTGCTTGCAGTGCCTTACCCTGCGGGAATCTTGCTCTGTT 343
QY 265 CTGGCTGGGGTCTGCTGGCGGAACGGGTGAGCTCAGCGGGTGTGTCTGCTGCTTCAAGGA 324
DB 344 CTGGCTGGGGTCTGCTGGCGGAACGGGTGAGCTCAGCGGGTGTGTCTGCTGCTTCAAGGA 403
QY 325 GGTCTCTGCCCCAGTGGCGGGGGCTGACCCAGAGCTTGCCTCCAGGAGAATGCTTAC 384
DB 404 GGTCTCTGCCCCAGTGGCGGGGGCTGACCCAGAGCTTGCCTCCAGGAGAATGCTTAC 463
QY 385 CGTGTCTGAGTGTGAGCTGTGCTGCTGTGAGAGGTCTGCANTAACTCTATGA 444
DB 464 CGTGTCTGAGTGTGAGCTGTGCTGCTGTGAGAGGTCTGCANTAACTCTATGA 523
QY 445 CCGGCTGTACACCCCAACATGTTCTGCGCGGGGAGGGCAAGACCAAGAGACTCCTG 504
DB 524 CCGGCTGTACACCCCAACATGTTCTGCGCGGGGAGGGCAAGACCAAGAGACTCCTG 583
QY 505 CAACG 509
DB 584 CAACG 588

RESULT 5
V11855
ID V11855 standard; cDNA: 1386 BP.
AC V11855;
DT 11-SEP-1998 (first entry)
DE Homo sapiens Tub Interactor (hTI-1) gene.
KW serine protease; tub interactor; treatment; obesity; cachexia;
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW neurodegenerative disease; Alzheimer's disease; drug screening;
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
OS Homo sapiens.
FH Key
FT CDS 2..701
FT CDS Location/Qualifiers
FT CDS /*tag= a

FT /product= hTI-1 protein
PN /note= "putative serine protease"
PD M09812302-A1.
PE 26-MAR-1998.
PF 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errada PR, Gimeno CJ;
DR WPI: 98-217246/19.
DR P-PSDB; W59129.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 10; Fig 1; 120pp; English.
CC The sequence is that of the Tub Interactor gene hTI-1 which
CC codes for a putative serine protease. TI genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 22.5%; Score 329; DB 1; Length 1386;
Best Local Similarity 83.1%; Pred. No. 2.3e-63;
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTACGCCGACACTGTTTCCAGAAGTGTGACAGAGCTCTTACACCATCGGGCTGGGCT 61
DB 100 GTACGCCGACACTGTTTCCAGAAGTGTGACAGAGCTCTTACACCATCGGGCTGGGCT 159
QY 62 GCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGTGAGGCCAGCTCTCCGT 121
DB 160 GCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGTGAGGCCAGCTCTCCGT 219
QY 122 ACGGACCCAGAGTACACAGACCCTTCTGCTTAAGACCTCATGCTCATCAAGTTGA 181
DB 220 ACGGACCCAGAGTACACAGACCCTTCTGCTTAAGACCTCATGCTCATCAAGTTGA 279
QY 182 CGAATCCGTGTCCAGTGTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCTAC 241
DB 280 CGAATCCGTGTCCAGTGTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCTAC 339
QY 242 CGCGGGGAATCTTGGCTCTGTTCTGGCTGGGGTCTGCTGGCGAAGCGGTGAGCTCACGGG 301
DB 340 CGCGGGGAATCTTGGCTCTGTTCTGGCTGGGGTCTGCTGGCGAAGCGGTGAGCTCACGGG 386
QY 302 TGTGTGTGCTGCTCTTCAAGGAGGTCTCTGCCAGTCCGCGGGGGCTGACCCAGAGCTC 361
DB 386 -----
QY 362 TGCCTCCAGCAGATGCTTACCGTGTGAGTGCAGTGCAGTGAAGTGTGCTGTGAN 421
DB 386 -----GGCAGATGCTTACCGTGTGAGTGCAGTGCAGTGAAGTGTGCTGTGAN 436
QY 422 GAGGTCTGCANTAACTCTATGACCCCGCTGTACACCCCAACATGTTCTGCGCGGGCGGA 481
DB 437 GAGGTCTGCANTAACTCTATGACCCCGCTGTACACCCCAACATGTTCTGCGCGGGCGGA 496
QY 482 GGGCAAGACCAAGAGACTCTGTCAACG 509
DB 497 GGGCAAGACCAAGAGACTCTGTCAACG 524

RESULT 6

V58644
ID V58644 standard; cDNA; 1248 BP.
AC V58644;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE1.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 217..696
FT /tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69387.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 112; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;
Best Local Similarity 83.1%; Pred. No. 2.2e-63;
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTCAGCCGCACACTGTTCCAGAGTGAAGTGCAGAGCTCCTACACCATCGGGCTGGCCCT 61
DB 120 GTCAGCCGCACACTGTTCCAGAGTGAAGTGCAGAGCTCCTACACCATCGGGCTGGCCCT 179
QY 62 GCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGT 121
DB 180 GCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGT 239
QY 122 ACGGACCCAGAGTACAACAGACCCTTGCTGCTAAGCACTCATAGCTCATCAAGTTGA 181
DB 240 ACGGACCCAGAGTACAACAGACCCTTGCTGCTAAGCACTCATAGCTCATCAAGTTGA 299
QY 182 CGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGGCCCTAC 241
DB 300 CGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGGCCCTAC 359
QY 242 CGCGGGGAAGTCTTGCCCTGTTCTTGCGCTGGGGTCTGCTGGCGAAGGTTGAGCTACGGG 301
DB 360 CGCGGGGAAGTCTTGCCCTGTTCTTGCGCTGGGGTCTGCTGGCGAAC----- 406
QY 302 TGTGTGTGCTGCTCTTCAAGAGGAGTCTCTGCGGAGTGGCGGGGCTGACCCAGAGCTC 361
DB 406 ----- 406
QY 362 TGCCTCCAGCAGATGCTACCGTGTGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 421
DB 406 -----GGCAGATGCTACCGTGTGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 456
QY 422 GAGGTCTGCANTAGCTCTATAGACCCGCTGTACCAACCCANCAATGTTCTGCGCCGGCGGA 481
DB 457 GAGGTCTGCAGTACTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGA 516
QY 482 GGGCAAGACCAGAGGACTCTCTGCAACG 509

Db 517 GGGCAAGACCAGAGGACTCTCTGCAACG 544

RESULT 7
ID V61249 standard; cDNA; 1248 BP.
AC V61249;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE1.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;
Best Local Similarity 83.1%; Pred. No. 2.2e-63;
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTCAGCCGCACACTGTTCCAGAGTGAAGTGCAGAGCTCCTACACCATCGGGCTGGCCCT 61
DB 120 GTCAGCCGCACACTGTTCCAGAGTGAAGTGCAGAGCTCCTACACCATCGGGCTGGCCCT 179
QY 62 GCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGT 121
DB 180 GCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGT 239
QY 122 ACGGACCCAGAGTACAACAGACCCTTGCTGCTAAGCACTCATAGCTCATCAAGTTGA 181
DB 240 ACGGACCCAGAGTACAACAGACCCTTGCTGCTAAGCACTCATAGCTCATCAAGTTGA 299
QY 182 CGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGGCCCTAC 241
DB 300 CGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGGCCCTAC 359
QY 242 CGCGGGGAAGTCTTGCCCTGTTCTTGCGCTGGGGTCTGCTGGCGAAGGTTGAGCTACGGG 301
DB 360 CGCGGGGAAGTCTTGCCCTGTTCTTGCGCTGGGGTCTGCTGGCGAAC----- 406
QY 302 TGTGTGTGCTGCTCTTCAAGAGGAGTCTCTGCGGAGTGGCGGGGCTGACCCAGAGCTC 361
DB 406 ----- 406
QY 362 TGCCTCCAGCAGATGCTACCGTGTGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 421
DB 406 -----GGCAGATGCTACCGTGTGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 456
QY 422 GAGGTCTGCANTAGCTCTATAGACCCGCTGTACCAACCCANCAATGTTCTGCGCCGGCGGA 481
DB 457 GAGGTCTGCAGTACTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGA 516
QY 482 GGGCAAGACCAGAGGACTCTCTGCAACG 509
DB 517 GGGCAAGACCAGAGGACTCTCTGCAACG 544


```
RESULT      8
V58647      8
ID V58647 standard; cDNA; 1167 BP.
AC V58647;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key
FT CDS      Location/Qualifiers
FT CDS      28..645
FT CDS      /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;
```

```
Query Match      20.1%; Score 292.8; DB 1; Length 1167;
Best Local Similarity 80.4%; Pred. No. 1.8e-55;
Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 25 AGTAGTCAGAGCTCCTACACATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 92 ACTGTTCCAGAACTCCTACACATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 151

QY 85 AGCCAGGAGCCAGATGTTGAGAGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 152 AGCCAGGAGCCAGATGTTGAGAGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 211

QY 145 CCTGCTCGCTAACGACCTCATCATCAAGTTGAGCAATCCGTGTCGAGTCTGACA 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 212 TCTTGCTCGCTAACGACCTCATCATCAAGTTGAGCAATCCGTGTCGAGTCTGACA 271

QY 205 CCATCCGAGCATCAGCATTTGCTTCGAGTGGCCCTACCCGGGGGAACCTTTGCCCTGTTT 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 CCATCCGAGCATCAGCATTTGCTTCGAGTGGCCCTACCCGGGGGAACCTTTGCCCTGTTT 331

QY 265 CTGGCTGGGGTCTGCTGGCGAAGCGGTGAGCTCAGCGGGTGTGTCTGCCCTCTTCAAGGA 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 332 CTGGCTGGGGTCTGCTGGCGAAC----- 355

QY 325 GGTCTCTGCCAGTGGGGGGGTGACCCAGAGCTGTGCTCCAGCAGAAATGCCTAC 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 -----GGCAGAAATGCCTAC 368

QY 385 CGTGTGAGTGCAGTGAACGTGTCGGTGTGTCTGANGAGGTCTGCANTAAAGTCTATGA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 CGTGTGAGTGCAGTGAACGTGTCGGTGTGTCTGANGAGTCTGCAGTAAGTCTATGA 428

QY 445 CCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGAGGCAAGACCAAGAGACTCCTG 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 429 CCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGAGGCAAGACCAAGAGACTCCTG 488
QY 505 CAACG 509
   ||| |||
DB 489 CAACG 493

RESULT      9
V61252
ID V61252 standard; cDNA; 1167 BP.
AC V61252;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
FH Key
FT CDS      Location/Qualifiers
FT CDS      27..645
FT CDS      /*tag= a
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;
```

```
Query Match      20.1%; Score 292.8; DB 1; Length 1167;
Best Local Similarity 80.4%; Pred. No. 1.8e-55;
Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 25 AGTAGTCAGAGCTCCTACACATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 92 ACTGTTCCAGAACTCCTACACATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG 151

QY 85 AGCCAGGAGCCAGATGTTGAGAGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 152 AGCCAGGAGCCAGATGTTGAGAGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 211

QY 145 CCTGCTCGCTAACGACCTCATCATCAAGTTGAGCAATCCGTGTCGAGTCTGACA 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 212 TCTTGCTCGCTAACGACCTCATCATCAAGTTGAGCAATCCGTGTCGAGTCTGACA 271

QY 205 CCATCCGAGCATCAGCATTTGCTTCGAGTGGCCCTACCCGGGGGAACCTTTGCCCTGTTT 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 CCATCCGAGCATCAGCATTTGCTTCGAGTGGCCCTACCCGGGGGAACCTTTGCCCTGTTT 331

QY 265 CTGGCTGGGGTCTGCTGGCGAAGCGGTGAGCTCAGCGGGTGTGTCTGCCCTCTTCAAGGA 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 332 CTGGCTGGGGTCTGCTGGCGAAC----- 355

QY 325 GGTCTCTGCCAGTGGGGGGGTGACCCAGAGCTTGTGCTCCAGCAGAAATGCCTAC 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 -----GGCAGAAATGCCTAC 368

QY 385 CGTGTGAGTGCAGTGAACGTGTCGGTGTGTCTGANGAGGTCTGCANTAAAGTCTATGA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 CGTGTGAGTGCAGTGAACGTGTCGGTGTGTCTGANGAGTCTGCAGTAAGTCTATGA 428

QY 445 CCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGAGGCAAGACCAAGAGACTCCTG 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 CCCGCTGTACCAACCCAGCATGTTCTGCGCCGGCGGAGGCAAGACCAAGAGACTCCTG 488
```


QY 505 CAACG 509
Db 489 CAACG 493

RESULT 10
V37495
ID V37495 standard; DNA; 871 BP.
AC V37495;
DT 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
OS benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
FH Homo sapiens.
Key Location/Qualifiers
FT CDS 31..777
FT /tag- a
FT /transl_except- (pos:367..369, aa:Xaa)
FT /transl_except- (pos:412..414, aa:Xaa)
FT /transl_except- (pos:424..426, aa:Xaa)
FT /product- "HPSK protein"
FT /note- "Xaa = unknown"

PN WO9820117-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goli SK;
DR WPI: 98-286933/25.
DR P-PSDB; W60592.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 5; Fig 1A-C; 68pp; English.
CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

Query Match 20.0%; Score 291.8; DB 1; Length 871;
Best Local Similarity 80.2%; Pred. No. 2.7e-55;
Matches 389; Conservative 0; Mismatches 13; Indels 83; Gaps 1;

QY 25 AGTAGTGACAGAGCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 84
Db 224 ACTGTTCCAGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 283
QY 85 AGCCAGGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGGACCCAGAGTACAACAGAC 144
Db 284 AGCCAGGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGGACCCAGAGTACAACAGAC 343

QY 145 CCTTGCTCGCTAACGACCTCATGCTCATCAAGTGGAGCAATCCGTCGAGTCTGACA 204
Db 344 CCTTGCTCGCTAACGACCTCATGTCATCAAGTGGAGCAATCCGTCGAGTCTGACA 403
QY 205 CCATCCGGAGCATGATGCTTCGACATGCCCTACCGGGGGAAGTCTTGCCTCGTT 264
Db 404 ACATCCGGAGNATCAGCATGNTTCGACATGCCCTACCGGGGGAAGTCTTGCCTCGTT 463
QY 265 CTGGCTGGGGTCTGCTGGCGGAACGGGTGAGCTCAGGGGTGTGTCTGCCCTCTTCAAGGA 324
Db 464 CTGGCTGGGGTCTGCTGGCGGAAC----- 487
QY 325 GGTCTCTGCCCCAGTCGGGGGGGCTGACCAGAGCTCTGCGTCCAGGCGAATGCTTAC 384
|||||

Db 487 -----GGCAGATGCCTAC 500
QY 385 CGTGCTGACGTGGGTGAACGTGTGCGTGGTGTCTGANGAGGTCGTGACANTAGCTATGA 444
Db 501 CGTGCTGACGTGGGTGAACGTGTGCGTGGTGTCTGAGGAGGTCTGAGTAAGCTATGA 560
QY 445 CCCGCTGACCAACCCANCATGTTCTGCGCCCGCGGAGGGCAAGACCAGAGACTCTTG 504
Db 561 CCCGCTGACCAACCCAGCATGTTCTGCGCCCGCGGAGGGCAAGACCAGAGACTCTTG 620
QY 505 CAACG 509
Db 621 CAACG 625

RESULT 11
X41114
ID X41114 standard; cDNA; 402 BP.
AC X41114;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:58.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR P-PSDB; Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulation
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 19.8%; Score 289.4; DB 1; Length 402;
Best Local Similarity 99.0%; Pred. No. 7.6e-55;
Matches 290; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTCAGCCGCACACTGTTTCCAGAGTGTGACAGAGCTCTCAACCATCGGGCTGGCCT 61
Db 108 GTCAGCCGCACACTGTTTCCAGAGTGTGATKACAGAGCTCTCAACCATCGGGCTGGCCT 167
QY 62 GCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGGCTTCGGT 121
Db 168 GCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGGCTTCGGT 227
QY 122 ACGGACCCAGAGTACAACAGAGACCTTGTCTGCTAACGACCTCATGCTCATCAATTGA 181
|||||

|||||
Db 228 ACGGCACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGA 287
QY 182 CGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATACAGATTGCTTCGACAGTGCCTAC 241
Db 288 CGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATACAGATTGCTTCGACAGTGCCTAC 347
QY 242 CGCGGGAACTCTTGCTCGTTCTGCTGGGCTGCTGCTGCGGAACGGTGAGC 294
Db 348 CGCGGGAACTCTTGCTCGTTCTGCTGGGCTGCTGCTGCGGAACGGCCAGC 400

RESULT 12

V58648 standard; cDNA; 1119 BP.
V58648;
AC V58648;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT CDS /*tag= a
PN MO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69389.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 116-117; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTAGTCAGAGCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 84
Db 98 ACTGTTCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 157
QY 85 AGCCAGGAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 144
Db 158 AGCCAGGAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 217
QY 145 CCTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCTCCGAGTGTGACA 204
Db 218 CCTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCTCCGAGTGTGACA 277
QY 205 CCATCCGAGCATCAGCATGCTTCGACAGTGCCCTACCGGGGAACTCTTGCTCGTTT 264
Db 278 CCATCCGAGCATCAGCATGCTTCGACAGTGCCCTACCGGGGAACTCTTGCTCGTTT 337
QY 265 CTGGCTGGGCTCTGCTGGCGAAGGTTG 291
Db 338 CTGGCTGGGCTCTGCTGGCGAAGCATG 364

RESULT 13
V61253
ID V61253 standard; cDNA; 1119 BP.
AC V61253;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN MO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 108-109; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTAGTCAGAGCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 84
Db 98 ACTGTTCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 157
QY 85 AGCCAGGAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 144
Db 158 AGCCAGGAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 217
QY 145 CCTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCTCCGAGTGTGACA 204
Db 218 CCTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCTCCGAGTGTGACA 277
QY 205 CCATCCGAGCATCAGCATGCTTCGACAGTGCCCTACCGGGGAACTCTTGCTCGTTT 264
Db 278 CCATCCGAGCATCAGCATGCTTCGACAGTGCCCTACCGGGGAACTCTTGCTCGTTT 337
QY 265 CTGGCTGGGCTCTGCTGGCGAAGGTTG 291
Db 338 CTGGCTGGGCTCTGCTGGCGAAGCATG 364

RESULT 14

T18325/c
ID T18325 standard; DNA; 24026 BP.
AC T18325;
DT 05-JUN-1996 (first entry)
DE BRCA1, human breast and ovarian cancer predisposing gene.
KW BRCA1; breast cancer; ovary cancer; predisposing gene;
KW susceptibility gene; diagnosis; prognosis; gene therapy; ds.
OS Homo sapiens.

FH Key Location/Qualifiers
FT Intron 1..55
FT Intron /*tag= a
FT Exon 56..155
FT Intron /*tag= b
FT Intron 156..1512
FT Exon /*tag= c
FT Exon 1513..1611

| | | | | | | | |
|----|--------|-------------|--|----|--------------|-------------|--|
| FT | intron | 1612.2206 | /*tag= d | FT | intron | 17727.18416 | /*tag= ae |
| FT | | 1612.2206 | /*tag= e | FT | | 17727.18416 | /*tag= ae |
| FT | exon | 2207.2260 | /note= "n at 1925-1937 represent an indefinite interval within the intron" | FT | exon | 18417.18787 | /note= "n at 18299-18312 represent an indefinite interval within the intron" |
| FT | intron | 2261.2677 | /*tag= f | FT | intron | 18788.19298 | /*tag= af |
| FT | | 2261.2677 | /*tag= g | FT | | 18788.19298 | /*tag= af |
| FT | exon | 2678.2788 | /note= "n at 2569-2581 represent an indefinite interval within the intron" | FT | exon | 19299.19386 | /note= "n at 18952-18964 represent an indefinite interval within the intron" |
| FT | intron | 2789.3328 | /*tag= h | FT | intron | 19387.20190 | /*tag= ah |
| FT | | 2789.3328 | /*tag= i | FT | | 19387.20190 | /*tag= ah |
| FT | exon | 3329.3406 | /note= "n at 3063-3075 represent an indefinite interval within the intron" | FT | exon | 20191.20267 | /note= "n at 19887-19899 represent an indefinite interval within the intron" |
| FT | intron | 3407.3813 | /*tag= j | FT | intron | 20268.21094 | /*tag= aj |
| FT | | 3407.3813 | /*tag= k | FT | | 20268.21094 | /*tag= aj |
| FT | exon | 3814.3902 | /note= "n at 3598-3610 represent an indefinite interval within the intron" | FT | exon | 21095.21135 | /note= "n at 20767-20779 represent an indefinite interval within the intron" |
| FT | intron | 3903.4224 | /*tag= l | FT | intron | 21136.21583 | /*tag= al |
| FT | | 3903.4224 | /*tag= m | FT | | 21136.21583 | /*tag= al |
| FT | exon | 4225.4364 | /note= "n at 4076-4088 represent an indefinite interval within the intron" | FT | exon | 21668.22233 | /note= "n at 21341-21353 represent an indefinite interval within the intron" |
| FT | intron | 4365.6571 | /*tag= n | FT | intron | 22234.22288 | /*tag= an |
| FT | | 4365.6571 | /*tag= o | FT | | 22234.22288 | /*tag= an |
| FT | exon | 6572.6677 | /note= "n at 4602-4614 represent an indefinite interval within the intron" | FT | exon | 22839.22832 | /note= "n at 21921-21933 represent an indefinite interval within the intron" |
| FT | intron | 6678.9163 | /*tag= p | FT | intron | 22907.23287 | /*tag= ap |
| FT | | 6678.9163 | /*tag= q | FT | | 22907.23287 | /*tag= ap |
| FT | exon | 9164.9207 | /*tag= r | FT | exon | 23288.23348 | /note= "n at 22567-22579 represent an indefinite interval within the intron" |
| FT | intron | 9208.10530 | /*tag= s | FT | intron | 23349.23698 | /*tag= ar |
| FT | exon | 10531.10607 | /*tag= t | FT | exon | 23699.24026 | /*tag= as |
| FT | intron | 10608.11597 | /*tag= u | FT | intron | 2725 | /note= "n at 23050-23062 represent an indefinite interval within the intron" |
| FT | | 10608.11597 | /*tag= v | FT | | 2725 | /*tag= au |
| FT | exon | 11598.15023 | /note= "n at 11383-11396 represent an indefinite interval within the intron" | FT | exon | 3653 | /note= "n at 23580-23592 represent an indefinite interval within the intron" |
| FT | intron | 15024.15424 | /*tag= w | FT | intron | 4391 | /*tag= at |
| FT | exon | 15425.15511 | /*tag= x | FT | exon | 4392 | /*tag= av |
| FT | intron | 15512.15952 | /*tag= y | FT | misc_feature | 6823 | /*tag= aw |
| FT | | 15512.15952 | /*tag= z | FT | | 6823 | /*tag= ax |
| FT | exon | 15953.16126 | /note= "n at 15647-15659 represent an indefinite interval within the intron" | FT | misc_feature | 9106 | /note= "polymorphic site" |
| FT | intron | 16127.16565 | /*tag= aa | FT | misc_feature | 9207 | /*tag= ay |
| FT | | 16127.16565 | /*tag= ab | FT | | 9207 | /*tag= az |
| FT | exon | 16566.16692 | /note= "n at 16370-16382 represent an indefinite interval within the intron" | FT | misc_feature | | /*tag= ba |
| FT | intron | 16693.17535 | /*tag= ac | FT | misc_feature | | /*tag= bb |
| FT | | 16693.17535 | /*tag= ad | FT | | | /*tag= bc |
| FT | exon | 17536.17726 | /note= "n at 17290-17302 represent an indefinite interval within the intron" | FT | misc_feature | | /*tag= bd |
| FT | | 17536.17726 | /*tag= ae | FT | | | /*tag= be |

| | | |
|----|--------------|--------------------|
| FT | /tag= | bc |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 9376 |
| FT | /tag= | bd |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 11908 |
| FT | /tag= | be |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 11994 |
| FT | /tag= | bf |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 12952 |
| FT | /tag= | bg |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13004 |
| FT | /tag= | bh |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13009 |
| FT | /tag= | bi |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13048 |
| FT | /tag= | bj |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13238 |
| FT | /tag= | bk |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13448 |
| FT | /tag= | bl |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13539 |
| FT | /tag= | bm |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 13951 |
| FT | /tag= | bn |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 14041 |
| FT | /tag= | bo |
| FT | /note= | "polymorphic site" |
| FT | misc_feature | 14046 |

Query Match 15.5%; Score 226.2; DB 1; Length 24026;
Best Local Similarity 87.0%; Pred. No. 1.3e-40;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

| | | | | |
|----|-------|---|---|------|
| QY | 1162 | GCTGGGCAGG | GTGCATGCCTGTATCCACACATTGGAGGC-GAGGCAGGCAGAT | 1220 |
| Db | 10357 | GCTGGGCACGGTGGCTCAAGCCTGTAATTCACACATTGGAGGCTGAGCGGGTGAC | 10298 | |
| QY | 1221 | CACCTGAGGTAAGGAGTTCAGACACCGCTGGCCAAATGGTGAATCCTGCTGACTA | 1280 | |
| Db | 10297 | CACCTGAGGTGATGAGTTCAGACACCGCTGGCCAAATGTTGAATCCCATCTCTACTA | 10238 | |
| QY | 1281 | AAATACAAAAGTTAGCTGATATGGTGGCAGCGCCTGTAATCCACGACTTGGAGG | 1340 | |
| Db | 10237 | AAATACAAAATTAGCGGGGTGTGGTGGCGCGCCTGTAGTCCAGCTACTAGAGAG | 10178 | |
| QY | 1341 | CTGAGGCAGGAGAAATGCTGAATATGGAGGCAGAGCTTGAAGTGAAGTGAATCAGAC | 1400 | |
| Db | 10177 | CTGAGGCAGGAGTATGCTTCAACCTGGAGGCAGACGTTGCGGAGAGAGATCAGAC | 10118 | |
| QY | 1401 | CACATATCTCCAGCTGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAA | 1459 | |
| Db | 10117 | CTCTGTACTCCAGCCTGGGCAACAGAGCAAGACTGTCTCAAAAAAAAAAAAAAAAA | 10059 | |

RESULT 15
T17455/c
ID T17455 standard; cDNA; 24025 BP.
AC T17455;
DT 07-OCT-1996 (first entry)
DE Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;

| | |
|----|--|
| KW | gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds. |
| OS | Homo sapiens. |
| FH | Key |
| FT | exon |
| FT | Location/Qualifiers |
| FT | 256..355 |
| FT | /tag= a |
| FT | /note= "exon 1" |
| FT | 356..1512 |
| FT | /tag= b |
| FT | /note= "intron 1" |
| FT | 1295 |
| FT | /tag= c |
| FT | /note= "known polymorphic site" |
| FT | 1513..1611 |
| FT | /tag= d |
| FT | /note= "exon 2" |
| FT | 1612..2206 |
| FT | /tag= e |
| FT | /note= "intron 2" |
| FT | 1925..1937 |
| FT | /tag= f |
| FT | /note= "indefinite interval within intron 2" |
| FT | 2141 |
| FT | /tag= g |
| FT | /note= "known polymorphic site" |
| FT | 2207..2260 |
| FT | /tag= h |
| FT | /note= "exon 3" |
| FT | 2261..2677 |
| FT | /tag= i |
| FT | /note= "intron 3" |
| FT | 2569..2581 |
| FT | /tag= j |
| FT | /note= "indefinite interval within intron 3" |
| FT | 2678..2788 |
| FT | /tag= k |
| FT | /note= "exon 4" |
| FT | 2725 |
| FT | /tag= l |
| FT | /note= "known polymorphic site" |
| FT | 2789..3328 |
| FT | /tag= m |
| FT | /note= "intron 4" |
| FT | 3063..3075 |
| FT | /tag= n |
| FT | /note= "indefinite interval within intron 4" |
| FT | 3329..3406 |
| FT | /tag= o |
| FT | /note= "exon 5" |
| FT | 3407..3813 |
| FT | /tag= p |
| FT | /note= "intron 5" |
| FT | 3598..3610 |
| FT | /tag= q |
| FT | /note= "indefinite interval within intron 5" |
| FT | 3653 |
| FT | /tag= r |
| FT | /note= "known polymorphic site" |
| FT | 3814..3902 |
| FT | /tag= s |
| FT | /note= "exon 6" |
| FT | 3903..4224 |
| FT | /tag= t |
| FT | /note= "intron 6" |
| FT | 4223 |
| FT | /tag= u |
| FT | /note= "site of 1 nucleotide deletion" |
| FT | 4076..4088 |
| FT | /tag= v |
| FT | /note= "indefinite interval within intron 6" |
| FT | 4225..4364 |
| FT | /tag= w |
| FT | /note= "exon 7" |
| FT | 4365..6571 |
| FT | intron |

FT /tag- x
FT /note- "intron 7"
FT 4391. .4392
FT /tag- y
FT /note- "known polymorphic site"
FT 4602. .4614
FT /tag- z
FT /note- "indefinite interval within intron 7"
FT 6538
FT /tag- aa
FT /note- "known polymorphic site"
FT 6572. .6677
FT /tag- ab
FT /note- "exon 8"
FT 6678. .9163
FT /tag- ac
FT /note- "intron 8"
FT 6823
FT /tag- ad
FT /note- "known polymorphic site"
FT 9106
FT /tag- ae
FT /note- "known polymorphic site"
FT 9164. .9209
FT /tag- af
FT /note- "exon 9"
FT 9207
FT /tag- ag
FT /note- "known polymorphic site"
FT 9210. .10530
FT /tag- ah
FT /note- "intron 9"
FT 9376
FT /tag- ai
FT /note- "known polymorphic site"
FT 10531. .10607
FT /tag- aj
FT /note- "exon 10"
FT 10608. .11597
FT /tag- ak
FT /note- "intron 10"
FT 11384. .11396
FT /tag- al
FT /note- "indefinite interval within intron 10"
FT 11598. .15023
FT /tag- am
FT /note- "exon 11"
FT 11908
FT /tag- an
FT /note- "known polymorphic site"
FT 11994
FT /tag- ao
FT /note- "known polymorphic site"
FT 12952
FT /tag- ap
FT /note- "known polymorphic site"
FT 13004
FT /tag- aq
FT /note- "known polymorphic site"
FT 13009
FT /tag- ar
FT /note- "known polymorphic site"
FT 13048
FT /tag- as
FT /note- "known polymorphic site"
FT 13238
FT /tag- at
FT /note- "known polymorphic site"
FT 13448
FT /tag- au
FT /note- "known polymorphic site"
FT 13539
FT /tag- av

FT /note- "known polymorphic site"
FT 13951
FT /tag- aw
FT /note- "known polymorphic site"
FT 14041
FT /tag- ax
FT /note- "known polymorphic site"
FT 14046
FT /tag- ay
FT /note- "known polymorphic site"
FT 14475
FT /tag- az
FT /note- "known polymorphic site"
FT 14874
FT /tag- ba
FT /note- "known polymorphic site"
FT 14891
FT /tag- bb
FT /note- "known polymorphic site"
FT 14966
FT /tag- bc
FT /note- "known polymorphic site"
FT 15024. .15424
FT /tag- bd
FT /note- "intron 11"
FT 15284
FT /tag- be
FT /note- "known polymorphic site"
FT 15425. .15511
FT /tag- bf
FT /note- "exon 12"
FT 15512. .15952
FT /tag- bg
FT /note- "intron 12"
FT 15647. .15659
FT /tag- bh
FT /note- "indefinite interval within intron 12"
FT 15953. .16126
FT /tag- bi
FT /note- "exon 13"
FT 16077
FT /tag- bj
FT /note- "known polymorphic site"
FT 16127. .16565
FT /tag- bk
FT /note- "intron 13"
FT 16243
FT /tag- bl

Query Match 15.5%; Score 226.2; DB 1; Length 24025;
Best Local Similarity 87.0%; Pred. No. 1.3e-40;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1162 GCTGGCAGGGTGCTCATGCGCTGTAATCCACGACTTTGGAGGC-GAGGCAGGCAGAT 1220
DB 10356 GCTGGCAGCGTGCTCAAGCCTGTAATCCACGACTTTGGAGGCTGAGCGGGTGAC 10297
QY 1221 CACTGAGGTAAAGAGTTCAGACCGCTGGCCAAAATGTAATCTGTCGTACTA 1280
DB 10296 CACTGAGGTCAAGTTCAGACCGCTGGCCAAACATGTTGAACCCCATCTACTA 10237
QY 1281 AAAATACAAAGTTAGCTGATATGTTGGCAGCGCGCTGTAATCCAGCTACTTGGAGG 1340
DB 10236 AAAATACAAATAGCCGGGTGTTGGCGCGCGCTGTAATCCAGCTACTAGAGAG 10177
QY 1341 CTGAGCAGAGAAATGCTTGAATATGGGAGGACAGAGTTGAAGTGAGTGAATCACAC 1400
DB 10176 CTGAGCAGAGATATGCTTCAACCTGGGAGGACAGACGTTGCGGAGAGGTGAGATCACAC 10117
QY 1401 CACTATACCTCAGCTGGGCAACAGAGTAAGACTCTGTCCAAAAAAAAAAAAAAAAA 1459
DB 10116 CTCTGACTCCAGCCTGGGCAACAGAGCAAGACTCTGTCCAAAAAAAAAAAAAAAAA 10058

Wed Sep 29 14:27:25 1999

us-09-030-606-174_1.rng

Page 12

Search completed: September 25, 1999, 09:57:00
Job time: 8071 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:29:29 ; Search time 165.01 Seconds
(without alignments)
810.594 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTCCAGCCGACACTGTTTC.....TCAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS9_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | length | DB | ID | Description |
|------------|-------|-------------|--------|----|-------------------|---------------------|
| 1 | 291.8 | 20.0 | 871 | 3 | US-08-744-026-2 | Sequence 2, Appl1 |
| c | 226.2 | 15.5 | 6769 | 2 | US-08-480-784-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 2 | US-08-483-553-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 2 | US-08-487-002-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 3 | US-08-483-554B-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 3 | US-08-488-011B-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 5 | PCT-US95-10202-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 5 | PCT-US95-10203-20 | Sequence 20, Appl1 |
| c | 226.2 | 15.5 | 6769 | 5 | PCT-US95-10220-20 | Sequence 20, Appl1 |
| c | 219.2 | 15.0 | 7676 | 3 | US-08-451-777A-7 | Sequence 7, Appl1 |
| c | 219.2 | 15.0 | 7676 | 4 | US-08-451-778A-7 | Sequence 7, Appl1 |
| c | 219.2 | 15.0 | 7676 | 4 | US-08-998-208-7 | Sequence 7, Appl1 |
| c | 219.2 | 15.0 | 7676 | 5 | PCT-US95-06743-7 | Sequence 7, Appl1 |
| c | 213.4 | 14.6 | 26764 | 1 | US-08-370-975B-1 | Sequence 1, Appl1 |
| c | 213.4 | 14.6 | 20303 | 1 | US-08-370-975B-6 | Sequence 1, Appl1 |
| c | 212.2 | 14.5 | 4823 | 5 | PCT-US94-08806-28 | Sequence 28, Appl1 |
| c | 212.2 | 14.5 | 4823 | 5 | PCT-US95-01829-5 | Sequence 5, Appl1 |
| c | 212.2 | 14.5 | 4823 | 5 | PCT-US95-16626-5 | Sequence 5, Appl1 |
| c | 211.8 | 14.5 | 22481 | 5 | PCT-US95-07201-43 | Sequence 43, Appl1 |
| c | 211.2 | 14.5 | 246240 | 4 | US-08-724-394A-20 | Sequence 20, Appl1 |
| c | 211.2 | 14.5 | 246240 | 4 | US-08-724-394A-21 | Sequence 21, Appl1 |
| c | 210.6 | 14.4 | 17327 | 1 | US-07-906-871-15 | Sequence 15, Appl1 |
| c | 209.8 | 14.4 | 2461 | 3 | US-08-832-883-3 | Sequence 3, Appl1 |
| c | 209.8 | 14.4 | 2461 | 4 | US-08-832-877-113 | Sequence 113, Appl1 |
| c | 209.8 | 14.3 | 1811 | 3 | US-08-848-252-1 | Sequence 1, Appl1 |
| c | 207.8 | 14.2 | 1988 | 4 | US-08-257-963B-11 | Sequence 11, Appl1 |
| c | 207.8 | 14.2 | 1988 | 5 | PCT-US95-07201-11 | Sequence 11, Appl1 |
| c | 207.6 | 14.2 | 7210 | 4 | US-08-257-963B-10 | Sequence 10, Appl1 |
| c | 207.6 | 14.2 | 7210 | 5 | PCT-US95-07201-10 | Sequence 10, Appl1 |
| c | 206.8 | 14.2 | 282 | 1 | US-08-133-629-8 | Sequence 8, Appl1 |
| c | 203.6 | 14.0 | 2477 | 2 | US-08-429-742-1 | Sequence 1, Appl1 |
| c | 203.2 | 13.9 | 7620 | 1 | US-07-767-135-1 | Sequence 1, Appl1 |
| c | 203.2 | 13.9 | 7620 | 1 | US-07-841-652-1 | Sequence 1, Appl1 |
| c | 202.4 | 13.9 | 246240 | 4 | US-08-724-394A-20 | Sequence 20, Appl1 |
| c | 202.4 | 13.9 | 246240 | 4 | US-08-724-394A-21 | Sequence 21, Appl1 |

| | | | | | | | |
|----|-------|-------|--------|-------|-------------------|--------------------|-------------------|
| 38 | 202.4 | 13.9 | 246240 | 4 | US-08-724-394A-22 | Sequence 22, Appl1 | |
| 39 | 201.4 | 13.8 | 3176 | 4 | US-08-910-733-17 | Sequence 17, Appl1 | |
| 40 | 199.4 | 13.7 | 8392 | 1 | US-08-080-255-6 | Sequence 6, Appl1 | |
| 41 | 199.4 | 13.7 | 8392 | 5 | PCT-US93-05857-6 | Sequence 6, Appl1 | |
| c | 42 | 198.4 | 13.6 | 8174 | 1 | US-07-914-281-5 | Sequence 5, Appl1 |
| c | 43 | 198.4 | 13.6 | 8174 | 1 | US-08-393-246-5 | Sequence 5, Appl1 |
| c | 44 | 198.4 | 13.6 | 11531 | 1 | US-08-068-945A-1 | Sequence 1, Appl1 |
| c | 45 | 198.4 | 13.6 | 3373 | 1 | US-08-273-411-2 | Sequence 2, Appl1 |

ALIGNMENTS

RESULT 1
US-08-744-026-2
; Sequence 2, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-744-026-2

Query Match 20.0%; Score 291.8; DB 3; Length 871;
Best Local Similarity 80.2%; Pred. No. 1.5e-64;
Matches 389; Conservative 0; Mismatches 13; Indels 83; Gaps 1;

| | | | |
|----|-----|--|-----|
| OY | 25 | AGTGAGTCAGAGCTCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG | 84 |
| Db | 224 | ACTGTTTCAGAACTCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAG | 283 |
| OY | 85 | AGCCAGGAGCCAGATGTGGAGGCGAGCCCTCTCCGTACGGCACCAGAGTACACAGAC | 144 |
| Db | 284 | AGCCAGGAGCCAGATGTGGAGGCGAGCCCTCTCCGTACGGCACCAGAGTACACAGAC | 343 |


```
QY 145 CCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACA 204
      |||
Db 344 CCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACA 403
QY 205 CCATCCGAGCATGACATGCTGCTGCGAGTGCCTTACCAGGGAACCTCTGCTGCTT 264
      |||
Db 404 ACATCCGAGNATGACATGCTGCTGCGAGTGCCTTACCAGGGAACCTCTGCTGCTT 463
QY 265 CTGGCTGGGCTGCTGGCGAAGCGTGAGCTCACGGGTGTGTGCTGCTGCTTCAAGGA 324
      |||
Db 464 CTGGCTGGGCTGCTGGCGAAGCGTGAGCTCACGGGTGTGTGCTGCTGCTTCAAGGA 487
QY 325 GGTCTCTGCCCCAGTCCGCGGGGCTGACCCAGAGCTCTGCTCCAGCAGATGCTTAC 384
      |||
Db 487 -----GGCAGATGCTTAC 500
QY 385 CGTGCTGACGTGCTGACGCTGTGCTGCTGTGANGAGGTCTGCANTAGCTCTATGA 444
      |||
Db 501 CGTGCTGACGTGCTGACGCTGTGCTGCTGTGANGAGGTCTGCANTAGCTCTATGA 560
QY 445 CCCGCTGTACACCCGANCATGTTCTGCGCGCGGAGGAGGCAAGACGAGACTCTCTG 504
      |||
Db 561 CCCGCTGTACACCCGAGCATGTTCTGCGCGCGGAGGAGGCAAGACGAGACTCTCTG 620
QY 505 CAACG 509
      |||
Db 621 CAACG 625
```

RESULT 2

```
US-08-480-784-20/c
; Sequence 20, Application US/08480784
; Patent No. 5693473
```

GENERAL INFORMATION:

```
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,784
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-480-784-20
```

```
Query Match 15.5%; Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
```

```
QY 1162 GCTGGCAGAGGTGGCTCATGCTGTAATCCAGCAGCTTTGGAGGC-GAGGACGACGAT 1220
      |||
Db 5743 GCTGGCAGAGGTGGCTCAAGCCTGTAATCCAGCAGCTTTGGAGGCCTGAGGCGGTGAC 5684
QY 1221 CACTTGAGGTAGAGGTTCAAGACGACCTGGCCAAATGCTGAATCCTGCTGACTA 1280
      |||
Db 5683 CACTTGAGGTAGAGTTCAAGACGACCTGGCCAAATGCTGAATCCTGCTGACTA 5624
QY 1281 AAATACAAAAGTTAGCTGATATGCTGGCAGCGCCCTGTAATCCAGCTACTTGGAGG 1340
      |||
Db 5623 AAATACAAAAGTTAGCTGATATGCTGGCAGCGCCCTGTAATCCAGCTACTTGGAGG 5564
QY 1341 CTGAGCAGAGCAATTCCTGTAATATGGGAGGAGGAGGTTGAAGTGAAGTGAATCAGAC 1400
      |||
Db 5563 CTGAGCAGAGCAATTCCTGTAATATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5504
QY 1401 CACTATACCTCAGCTGGGCAACAGAGTAACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
      |||
Db 5503 CTCTGTACTCAGCCTGGGCAACAGAGTAACTGTCTCAAAAAAAAAAAAAAAAAAAAA 5445
```

RESULT 3

```
US-08-483-553-20/c
; Sequence 20, Application US/08483553
; Patent No. 5709999
```

GENERAL INFORMATION:

```
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
```


STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEO ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-553-20

Query Match 15.5%; Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1162 GCTGGCAGGGTGGCTCATGCTGTAATCCAGCACTTTGGAGGC-GAGGCAGGCAGAT 1220
Db 5743 GCTGGCAGCGTGGCTCAAGCCTGTAATTCAGCACTTTGGAGGCGTGGAGCGGTGGAC 5684
QY 1221 CACTTGAGTAGAGTTCAAGACCAGCCTGGCCAAATGTTGAATCTGCTCTACTA 1280
Db 5683 CACTTGAGGTATGATTCAGACCAGCCTGGCCCAACATGTTGAATCCCACTCTACTA 5624
QY 1281 AAAATACAAAGTTAGCTGATATGTTGGCAGGCGCTGTAATCCAGCTACTTGGAGG 1340
Db 5623 AAAATACAAAGTTAGCGGGGTGTGTGGCGCGCCTGTAGTCCAGCTACTAGAGAGG 5564
QY 1341 CTGAGGCAGAGATTGCTGAATATGGGAGGAGAGGTTGAAGTGAAGTGAATGATCAGAC 1400
Db 5563 CTGAGGCAGAGATTGCTGAACCTGGAGGAGAGAGCTTGGAGAGGTGAGATCAGAC 5504
QY 1401 CACTATACTCCAGCTGGGCAACAGAGTAAGACTCTGCTCAAAAAAAAAAAAAAAAAA 1459
Db 5503 CTCTGACTCCAGCCTGGGCAACAGAGTAAGACTCTGCTCAAAAAAAAAAAAAAAAAA 5445

RESULT 4
US-08-487-002-20/c
Sequence 20, Application US/08487002
Patent No. 5710001
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEO ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-20

Query Match 15.5%; Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1162 GCTGGCAGGGTGGCTCATGCTGTAATCCAGCACTTTGGAGGC-GAGGCAGGCAGAT 1220
Db 5743 GCTGGCAGCGTGGCTCAAGCCTGTAATTCAGCACTTTGGAGGCGTGGAGCGGTGGAC 5684

| | | | | | | |
|----|------|-----|---------------------------|-----------------------|---------------------|------|
| QY | 1221 | CAC | TGAGGTAAGGAGTTCAAGACCAGCC | TGGCCAAAATG | TGTAATCCTGTCTGTACTA | 1280 |
| | | | | | | |
| Db | 5683 | CAC | TGAGGTATGAGTTCAAGACCAGCC | TGGCCAAACATG | TGTAATCCCATCTCTACTA | 5624 |
| QY | 1281 | AAA | TACAAAAGTTAGCTGATATG | TGTGGCAGGCGCCTGTAATCC | CAGCTACTTGGAGG | 1340 |
| | | | | | | |
| Db | 5623 | AAA | ATACAAAATTAGCCGGGTGTG | TGTGGCGCGCCTGTAGTCC | CAGCTACTAGAGAGG | 5564 |
| QY | 1341 | CTG | AGGCAGGAGATTCGTTGAATAT | TGGAGGCAGAGGTTGAAGT | GAGTTGAGATCAAC | 1400 |
| | | | | | | |
| Db | 5563 | CTG | AGGCAGGAGATATGCGCTTCA | ACCTGGGAGGCAGACGTTG | CGGAGAGGTGAGATCAAC | 5504 |
| QY | 1401 | CAC | TATACTCCAGCTGGGGCAAC | AGAGTAAGACTCTGTCTC | AAAAAAAAAAAAAAAA | 1459 |
| | | | | | | |
| Db | 5503 | CTC | TGTACTCCAGCCTGGGCAAC | AGAGCAAGACTCTGTCTC | AAAAAAAAAAAAAAAA | 5445 |

RESULT 5
US-08-483-554B-20/c
Sequence 20, Application US/08483554B
Patent No. 5747282
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eldens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,554B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300

```

; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 6769 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens
;
US-08-483-554B-20

```

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 15.5% | Score 226.2; | DB 3; | Length 6769; |
| Best Local Similarity | 87.0%; | Pred. No. 9.5e-48; | | |
| Matches 260; Conservative | 0; | Mismatches 38; | Indels 1; | Gaps 1; |

[illegible]

```

1  RESULT 6
2  US-08-488-011B-20/c
3  ; Sequence 20, Application US/08488011B
4  ; Patent No. 5753441
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Skolnick, Mark H.
9  ; APPLICANT: Goldgar, David E.
10 ; APPLICANT: Miki, Yoshio
11 ; APPLICANT: Swenson, Jeff
12 ; APPLICANT: Kamb, Alexander
13 ; APPLICANT: Harshman, Keith D.
14 ; APPLICANT: Shattuck-Eldens, Donna M.
15 ; APPLICANT: Tavtigian, Sean V.
16 ; APPLICANT: Wiseman, Roger W.
17 ; APPLICANT: Futreal, P. Andrew
18 ; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
19 ; TITLE OF INVENTION: Susceptibility Gene
20 ;
21 ; NUMBER OF SEQUENCES: 85
22 ;
23 ; CORRESPONDENCE ADDRESSES:
24 ;
25 ; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
26 ; STREET: 1201 New York Avenue, N.W., Suite 1000
27 ; CITY: Washington
28 ; STATE: DC
29 ;
30 ; COUNTRY: USA
31 ; ZIP: 20005
32 ;
33 ; COMPUTER READABLE FORM:
34 ; MEDIUM TYPE: Floppy disk
35 ;
36 ; COMPUTER: IBM PC compatible
37 ; OPERATING SYSTEM: PC-DOS/MS-DOS
38 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
39 ; CURRENT APPLICATION DATA:
40 ; APPLICATION NUMBER: US/08/488,011B
41 ; FILING DATE: 07-JUN-1995
42 ;
43 ; CLASSIFICATION: 435

```


| DB | 5563 | CTGAGGCAGGAGTATGGCTTCAACTGGGAGGCAGACGTTGGGAGAGGTGAGATCACAC | 5504 |
|----|------|---|------|
| QY | 1401 | CACATATACTCCAGCTGGGGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAAA | 1459 |
| Db | 5503 | CTCTGTACTCCAGCCTGGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAA | 5445 |

```

RESULT      6
PCT-US95-10203-20/c
; Sequence 20, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10203
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
PCT-US95-10203-20

```

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 15.5%; | Score 226.2; | DB 5; | Length 6769; |
| Best Local Similarity | 87.0%; | Pred. No. 9.5e-48; | | |
| Matches 260; | Conservative 0; | Mismatches 38; | Indels 1; | Gaps 1; |

| | | | |
|----|------|--|------|
| QY | 1162 | GCTGGGCAGGCGTGGCTCATGCCCTGTAAATCCACGACTTTGGGAGGC - GAGGCAGGCAGAT | 1220 |
| Db | 5743 | GCTGGGCACGGTGCTCAAGCCTGTAAATTCACGACTTTGGGAGGCTGAGGCGGGGTGGAC | 5684 |
| QY | 1221 | CACCTGAGGTAAGGAGCTTCAAGACCAGCCTGGGCCAAAATGGTGAATCCTGTCTGTA | 1280 |
| Db | 5683 | CACCTGAGGTCAITGAGTTCAGAAGACCAGCCTGGCCACAACATGGTGAACCCTCATCTACTA | 5624 |
| QY | 1281 | AAAAATACAAAAGTTAGCTGCATATGTGTGCAGGCGCCTGTAAATCCAGCTACTTGGGAGG | 1340 |
| Db | 5623 | AAAATACAAAAATTAGCCGGGTGTGTGTGCGCGCGCCTGTAGTCCCAGCTACTAGAGAGG | 5564 |
| QY | 1341 | CTGAGGCAGGAGAATTGCTTGAATATGGAGGCGCAGAGGTTGAAGTGAGTGAAGATCACAC | 1400 |
| Db | 5563 | CTGAGGCAGGAGTAGTATGCGCTTCAACCTGGGAGGCGCAGACGTTCCGGAGAGGTGAGATCACAC | 5504 |
| QY | 1401 | CACATACTCCAGCTGGGGCAACAGAGTAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA | 1459 |
| Db | 5503 | CTCTGTACTCCAGCCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA | 5445 |

RESULT 9
PCT-US95-10220-20/c
; Sequence 20, Application PC/TUS9510220
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eldens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: Method for Diagnosing a
; TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10220
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824


```

: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08-308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6769 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
:
PCT-US95-10220-20

```

```

Query Match      15.5%; Score 226.2; DB 5; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1162 GCTGGCGAGGGTGGCTCATGCTGTAAATCCAGCAGCACTTTGGAGGC-GAGGCAGGCAGAT 1220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5743 GCTGGCGACGGTGGCTCAAGCCTGTAAATCCAGCACTTTGGAGGCTGAGCGGGTGGAC 5684
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1221 CACTTGAGGTAAGAGATTCAAGACCAGCCCTGGCCAAATGTTGAATCCTGTCTACTA 1280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5683 CACTTGAGGTCATGAGTTCAGACCCAGCCCTGGCCAAACATGGTGAACCCCATCTCTACTA 5624
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1281 AAAATACAAAAGTAGCTGATATGCTGGCAGGCGCCCTGTAATCCAGCTACTTGGAGG 1340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5623 AAAATACAAAATTAGCCGGGTGTGTGGCGCGCCCTGTAGTCCAGCTACTAGAGAGG 5564
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1341 CTGAGCGCAGGAGAAATTGCTGAATATGGGAGGAGGAGGAGGTTGAAGTGAAGTTGAGATCACAC 1400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5563 CTGAGCGCAGGAGATGAGCTTCAACCTGGGAGGAGGAGCAGACGTTGCGGAGGAGGTGAGATCACAC 5504
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1401 CACTATACTCCAGCTGGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5503 CTCTGTACTCCAGCCTGGGCAACAGAGCAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 5445
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-08-451-777A-7/c
; Sequence 7, Application US/08451777A
; Patent No. 5789223
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Stambolian, Dwight
; TITLE OF INVENTION: Human Galactokinase Gene
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: 709 Swedeland Road/TW22220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/451,777A
: FILING DATE: 26-MAY-1995
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/10825
: FILING DATE: 23-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Eagle, Alissa M.
: REGISTRATION NUMBER: 37,126
: REFERENCE/DOCKET NUMBER: P50268-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5364
: TELEFAX: 610-270-5090
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-451-777A-7

```

```

Query Match      15.0%; Score 219.2; DB 3; Length 7676;
Best Local Similarity 85.3%; Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 1161 GGCTGGGCAGGGTGGCTCATGCTGTATGCCAGACACTTGGGAGGC-GAGGCAGGCAGA 1219
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6058 GGGTGGGCGCGGTGGCTTCACACCTGTATCCAGCACATTGGGAAGCTGAGGCAGGCAGA 5999

QY 1220 TCACTTGAGGTAAGAGATTCAAGACCAGCCCTGGCCAAAATGGTGAATCCTGTCTACT 1279
    | | ||||| ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5998 TTGCCTGAGGTACAGGAGCTCGAAGACCAGCCCTGGCCACATGGCGAAACCCGCTCTACT 5939

QY 1280 AAAAATACAAAAGTTAGCTGGATATGTTGGCAGGCGCCTGTATCCAGCTACTTGGGAG 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5938 AAAAATACAAAAGTTAGCGGGCATCGTGGCAGGCGCCTGTATCCAGCTACTCAGGAG 5879

QY 1340 GCTGAGGCAGAGAAATGCTTGAATATGGAGGCAGAGAGTTGAAGTGAGTTGAGATCACA 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5878 GCTGAGGCAGAGAAATGCTTGAACCTGGAGGCGGAGGTTGCAGTGAGTGAGATCATG 5819

QY 1400 CCACTATACTCCAGCTGGGGCAACAGAGTAAGACTGTGTCTCAAAAAAAAAAAAAAAAA 1459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5818 CCACTGCACCTCAGCCCTGGGTGACAGAGTGAGGCTCCCTCTCAAAAAAAAAAGAAAGA 5759

RESULT 11
US-08-451-778A-7/c
; Sequence 7, Application US/08451778A
; Patent No. 5830649
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Stambolian, Dwight
; TITLE OF INVENTION: Human Galactokinase Gene
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: 709 Swedeland Road/UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,778A
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10825
; FILING DATE: 23-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Eagle, Alissa M.
; REGISTRATION NUMBER: 37,126
; REFERENCE/DOCKET NUMBER: P50268-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5364
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-451-778A-7
```

```
Query Match 15.0%; Score 219.2; DB 4; Length 7676;
Best Local Similarity 85.3%; Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
```

```
QY 1161 GGCTGGGCGAGGTGGCTCATGCTGTATCCCGACACTTTGGAGGC-GAGGCGAGCAGA 1219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6058 GGGTGGGCGCGGTGGCTCACACCTGTATCCCGACACTTTGGAGGCTGAGGCGAGCAGA 5999

QY 1220 TCACTTGAGGTAGAGTTCAGACCAAGCCCTGGCCAAATGGTGAATCCTGTCTACT 1279
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5998 TTGCCTGAGGTAGAGCTCGAGACCAGCCCTGGCCAAACATGGCGAAACCCGCTCTACT 5939

QY 1280 AAAAATACAAAGTTAGCTGATATGTTGGCAGCGCCCTGTAATCCAGCTACTGGGAG 1339
   ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 5938 AAAAATACAAAGTTAGCCGGGCATCGTGGCAGCGCCCTGTAATCCAGCTACTCAGAG 5879

QY 1340 GCTGAGGCGAGGAATGCTGAATATGGGAGGAGAGGTTGAAGTGAAGTGAAGTCA 1399
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5878 GCTGAGGCGAGGAATGCTGAACCTGGGAGGCGAGGTTGCAGTGAAGTGAAGTCA 5819

QY 1400 CCACTATACTCCAGCTGGGCGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAA 1459
   ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 5818 CCACTGCACTTCAGCCTGGGTGACAGAGTGAAGGCTCCCTCTCAAAAAAAAAAAAAAGA 5759
```

```
RESULT 12
US-08-998-208-7/c
; Sequence 7, Application US/08998208
; Patent No. 5880105
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Stambolian, Dwight
; TITLE OF INVENTION: Human Galactokinase Gene
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: 709 Swedeland Road/UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,208
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/451,777
; FILING DATE: 26-MAY-1995
; APPLICATION NUMBER: PCT/US94/10825
; FILING DATE: 23-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Eagle, Alissa M.
; REGISTRATION NUMBER: 37,126
; REFERENCE/DOCKET NUMBER: P50268-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5364
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-998-208-7
```

```
Query Match 15.0%; Score 219.2; DB 4; Length 7676;
Best Local Similarity 85.3%; Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
```

```
QY 1161 GGCTGGGCGAGGTGGCTCATGCTGTATCCCGACACTTTGGAGGC-GAGGCGAGCAGA 1219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6058 GGGTGGGCGCGGTGGCTCACACCTGTATCCCGACACTTTGGAGGCTGAGGCGAGCAGA 5999

QY 1220 TCACTTGAGGTAGAGTTCAGACCAAGCCCTGGCCAAATGGTGAATCCTGTCTACT 1279
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5998 TTGCCTGAGGTAGAGCTCGAGACCAGCCCTGGCCAAACATGGCGAAACCCGCTCTACT 5939

QY 1280 AAAAATACAAAGTTAGCTGATATGTTGGCAGCGCCCTGTAATCCAGCTACTGGGAG 1339
   ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 5938 AAAAATACAAAGTTAGCCGGGCATCGTGGCAGCGCCCTGTAATCCAGCTACTCAGAG 5879

QY 1340 GCTGAGGCGAGGAATGCTGAATATGGGAGGAGGAGGTTGAAGTGAAGTGAAGTCA 1399
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5878 GCTGAGGCGAGGAATGCTGAACCTGGGAGGCGGAGGTTGCAGTGAAGTGAAGTCA 5819

QY 1400 CCACTATACTCCAGCTGGGCGCAACAGAGTAAGACTCTGTCTCAAAAAAAAAAAAA 1459
   ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 5818 CCACTGCACTTCAGCCTGGGTGACAGAGTGAAGGCTCCCTCTCAAAAAAAAAAAAAAGA 5759
```

```
RESULT 13
PCT-US95-06743-7/c
; Sequence 7, Application PC/TUS9506743
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Stambolian, Dwight
; TITLE OF INVENTION: Human Galactokinase Gene
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: 709 Swedeland Road/UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
```


FT Misc_difference 12 /note= "O-glycosylated"
FT Misc_difference 24 /note= "O-glycosylated"
FT Misc_difference 36 /note= "O-glycosylated"
FT Misc_difference 43 /note= "O-glycosylated"
FT Misc_difference 46 /note= "O-glycosylated"
FT Misc_difference 67 /note= "O-glycosylated"
FT Misc_difference 68 /note= "O-glycosylated"
FT Misc_difference 73 /note= "O-glycosylated"
FT Misc_difference 88 /note= "O-glycosylated"
FT Misc_difference 88 /note= "O-glycosylated"
PN W09733911-A1.
PD 18-SEP-1997.
PF 14-MAR-1997; U03461.
PR 14-MAR-1996; US-615944.
PA (UNIW) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI; 97-470818/43.
PI N-PSDB: T90761.
PT GDNF-neurturin family related growth factor, Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PT insufficiency
PS Claim 3; Page 151; 228pp; English.
PS This polypeptide comprises mature mouse persephin, a novel growth
CC factor and member of the glial-derived neurotrophic factor-
CC neurturin family. Recombinant mature persephin can be expressed
CC in host cells utilizing an isolated nucleic acid sequence (see
CC T90761). Persephin polypeptides, and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure. Persephin can be
CC added to a cell culture medium to promote growth and/or
CC differentiation.
SQ Sequence 96 AA;

```

Query Match          62.2%; Score 620; DB 26; Length 96;
Best Local Similarity 81.3%; Pred. No. 3,48e-53;
Matches      78; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

Db      1 alagscrlwsltlpvaelglgyaseekvlfrycagscpgcartqslvarlgrgrahg 60
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY      38 ALSGPCQLWSLTLVAELGLGYASEEKVIFRYCAGSCPRGARTQGLALARLQGGRAGH 97

Db      61 rpccqptsyadvtflddqhhwqglpqlsaacgcgg 96
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY      98 GPCCRPTRYTDVAFLLDDRHRWQRLPQLSAAACGCGG 133

RESULT      5
ID      W30068 standard; Protein; 91 AA.
AC      W30068;
DT      27-MAR-1998 (first entry)
DE      Mature rat persephin.
KW      Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
KW      neuronal degeneration; haematopoietic cell degeneration;
KW      cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW      neuropathy; Alzheimer's disease; Parkinson's disease; stroke;

```

KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; rat.
 OS Rattus sp.
 PN WO9733911-A1.
 PD 18-SEP-1997.
 PF 14-MAR-1997; U03461.
 PR 14-MAR-1996; US-615944.
 PA (UNIV) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI; 97-470818/43.
 PT GDNF-neurturin family related growth factor, Persephin - used to
 PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
 PT insufficiency
 PS Claim 3; Page 152; 228pp; English.
 PS This polypeptide comprises mature rat persephin, a novel growth
 CC factor and member of the glial-derived neurotrophic factor-
 CC neurturin family. Recombinant mature persephin can be expressed
 CC in host cells utilising an isolated nucleic acid sequence.
 CC Persephin polypeptides, and DNA sequences encoding them,
 CC can be used in claimed methods to prevent or treat cellular
 CC degeneration or insufficiency, such as neuronal degeneration
 CC resulting from peripheral neuropathy, amyotrophic lateral
 CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, ischaemic stroke, acute brain or spinal cord injury,
 CC nervous system tumours, multiple sclerosis and infection. The
 CC cellular degeneration or insufficiency may also comprise
 CC haematopoietic cell degeneration or insufficiency resulting from
 CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
 CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
 CC included is cardiac muscle degeneration or insufficiency arising
 CC from cardiomyopathy or congestive heart failure. Persephin can be
 CC added to a cell culture medium to promote growth and/or
 CC differentiation.
 CC Sequence 91 AA;
 SQ

[illegible]

RESULT 6
ID W30065 standard; Protein; 89 AA.
AC W30065;
DT 27-MAR-1998 (first entry)
DE Mature rat persephin.
KW Persephin; neurtrurin; glial-derived neurotrophic factor; GDNF;
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
KW Huntington's disease; nervous system tumour; multiple sclerosis;
KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
KW congestive heart failure; therapy; rat.
OS Rattus sp.
PN W09733911-AL.
PD 18-SEP-1997.
PE 14-MAR-1997; U03461.
PR 14-MAR-1996; US-615944.
PA (UNITW) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI: 97-470818/43.
DR N-PSDB; T90760.
PT GDNF-neurtrurin family related growth factor, Persephin - used to

PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PS Claim 2; page 152; 228pp: English.
CC This polypeptide comprises mature rat persephin, a novel growth
CC factor and member of the glial-derived neurotrophic factor-
CC neurturin family. Recombinant mature persephin can be expressed
CC in host cells utilising an isolated nucleic acid sequence (see
CC T90760). Persephin polypeptides, and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure. Persephin can be
CC added to a cell culture medium to promote growth and/or
CC differentiation.
SQ Sequence 89 AA;

| | | | | |
|-----------------------|------------------|---------------------|-----------|------------|
| Query Match | 58.78; | Score 585; | DB 26; | Length 89; |
| Best Local Similarity | 80.98; | Pred. No. 1.90e-49; | | |
| Matches | 72; Conservative | 12; Mismatches | 5; Indels | 0; Gaps 0; |

```
Db      1 crwsltlpwaelgyaseekifrfycagscpqevrtqtslvarlrvggrahyrpcqg 60  
       |:|||||:|||||||:|||||||: |||: | ||||:||||| |||:  
QY     43 cqlmsltlsvalglcyaseekvifrvcagscprrgartqhglalarkggcrahgppccr 102
```

```

Db      61  ptsyadvtfldhnhhwqglpqlsaacgc 89
        ||:|||||:|:|||||
QY     103  PTRYTDVAFLDDRHRWQRLPQLSAACGC 131

```

| | |
|--------|---------------------------------|
| RESULT | 7 |
| ID | W30064 standard; Protein; 89 AA |
| AC | W30064; |
| DT | 27-MAR-1998 (first entry) |

KW Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
KW Huntington's disease; nervous system tumour; multiple sclerosis;
KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
KW congestive heart failure; therapy; mouse.
OS Mus musculus.

| FH | key | Location/Qualifiers |
|-----|-----|---------------------|
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 7 | 7 |
| 8 | 8 | 8 |
| 9 | 9 | 9 |
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12 |
| 13 | 13 | 13 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |
| 21 | 21 | 21 |
| 22 | 22 | 22 |
| 23 | 23 | 23 |
| 24 | 24 | 24 |
| 25 | 25 | 25 |
| 26 | 26 | 26 |
| 27 | 27 | 27 |
| 28 | 28 | 28 |
| 29 | 29 | 29 |
| 30 | 30 | 30 |
| 31 | 31 | 31 |
| 32 | 32 | 32 |
| 33 | 33 | 33 |
| 34 | 34 | 34 |
| 35 | 35 | 35 |
| 36 | 36 | 36 |
| 37 | 37 | 37 |
| 38 | 38 | 38 |
| 39 | 39 | 39 |
| 40 | 40 | 40 |
| 41 | 41 | 41 |
| 42 | 42 | 42 |
| 43 | 43 | 43 |
| 44 | 44 | 44 |
| 45 | 45 | 45 |
| 46 | 46 | 46 |
| 47 | 47 | 47 |
| 48 | 48 | 48 |
| 49 | 49 | 49 |
| 50 | 50 | 50 |
| 51 | 51 | 51 |
| 52 | 52 | 52 |
| 53 | 53 | 53 |
| 54 | 54 | 54 |
| 55 | 55 | 55 |
| 56 | 56 | 56 |
| 57 | 57 | 57 |
| 58 | 58 | 58 |
| 59 | 59 | 59 |
| 60 | 60 | 60 |
| 61 | 61 | 61 |
| 62 | 62 | 62 |
| 63 | 63 | 63 |
| 64 | 64 | 64 |
| 65 | 65 | 65 |
| 66 | 66 | 66 |
| 67 | 67 | 67 |
| 68 | 68 | 68 |
| 69 | 69 | 69 |
| 70 | 70 | 70 |
| 71 | 71 | 71 |
| 72 | 72 | 72 |
| 73 | 73 | 73 |
| 74 | 74 | 74 |
| 75 | 75 | 75 |
| 76 | 76 | 76 |
| 77 | 77 | 77 |
| 78 | 78 | 78 |
| 79 | 79 | 79 |
| 80 | 80 | 80 |
| 81 | 81 | 81 |
| 82 | 82 | 82 |
| 83 | 83 | 83 |
| 84 | 84 | 84 |
| 85 | 85 | 85 |
| 86 | 86 | 86 |
| 87 | 87 | 87 |
| 88 | 88 | 88 |
| 89 | 89 | 89 |
| 90 | 90 | 90 |
| 91 | 91 | 91 |
| 92 | 92 | 92 |
| 93 | 93 | 93 |
| 94 | 94 | 94 |
| 95 | 95 | 95 |
| 96 | 96 | 96 |
| 97 | 97 | 97 |
| 98 | 98 | 98 |
| 99 | 99 | 99 |
| 100 | 100 | 100 |

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

CT 20180727T0-3CTW T3

MSC: 47D03, 47D07, 47D08, 47D10, 47D15, 47D20, 47D25, 47D30, 47D35, 47D40, 47D45, 47D50, 47D55, 47D60, 47D65, 47D70, 47D75, 47D80, 47D85, 47D90, 47D95, 47D100, 47D105, 47D110, 47D115, 47D120, 47D125, 47D130, 47D135, 47D140, 47D145, 47D150, 47D155, 47D160, 47D165, 47D170, 47D175, 47D180, 47D185, 47D190, 47D195, 47D200, 47D205, 47D210, 47D215, 47D220, 47D225, 47D230, 47D235, 47D240, 47D245, 47D250, 47D255, 47D260, 47D265, 47D270, 47D275, 47D280, 47D285, 47D290, 47D295, 47D300, 47D305, 47D310, 47D315, 47D320, 47D325, 47D330, 47D335, 47D340, 47D345, 47D350, 47D355, 47D360, 47D365, 47D370, 47D375, 47D380, 47D385, 47D390, 47D395, 47D400, 47D405, 47D410, 47D415, 47D420, 47D425, 47D430, 47D435, 47D440, 47D445, 47D450, 47D455, 47D460, 47D465, 47D470, 47D475, 47D480, 47D485, 47D490, 47D495, 47D500, 47D505, 47D510, 47D515, 47D520, 47D525, 47D530, 47D535, 47D540, 47D545, 47D550, 47D555, 47D560, 47D565, 47D570, 47D575, 47D580, 47D585, 47D590, 47D595, 47D600, 47D605, 47D610, 47D615, 47D620, 47D625, 47D630, 47D635, 47D640, 47D645, 47D650, 47D655, 47D660, 47D665, 47D670, 47D675, 47D680, 47D685, 47D690, 47D695, 47D700, 47D705, 47D710, 47D715, 47D720, 47D725, 47D730, 47D735, 47D740, 47D745, 47D750, 47D755, 47D760, 47D765, 47D770, 47D775, 47D780, 47D785, 47D790, 47D795, 47D800, 47D805, 47D810, 47D815, 47D820, 47D825, 47D830, 47D835, 47D840, 47D845, 47D850, 47D855, 47D860, 47D865, 47D870, 47D875, 47D880, 47D885, 47D890, 47D895, 47D900, 47D905, 47D910, 47D915, 47D920, 47D925, 47D930, 47D935, 47D940, 47D945, 47D950, 47D955, 47D960, 47D965, 47D970, 47D975, 47D980, 47D985, 47D990, 47D995, 47D1000, 47D1005, 47D1010, 47D1015, 47D1020, 47D1025, 47D1030, 47D1035, 47D1040, 47D1045, 47D1050, 47D1055, 47D1060, 47D1065, 47D1070, 47D1075, 47D1080, 47D1085, 47D1090, 47D1095, 47D1100, 47D1105, 47D1110, 47D1115, 47D1120, 47D1125, 47D1130, 47D1135, 47D1140, 47D1145, 47D1150, 47D1155, 47D1160, 47D1165, 47D1170, 47D1175, 47D1180, 47D1185, 47D1190, 47D1195, 47D1200, 47D1205, 47D1210, 47D1215, 47D1220, 47D1225, 47D1230, 47D1235, 47D1240, 47D1245, 47D1250, 47D1255, 47D1260, 47D1265, 47D1270, 47D1275, 47D1280, 47D1285, 47D1290, 47D1295, 47D1300, 47D1305, 47D1310, 47D1315, 47D1320, 47D1325, 47D1330, 47D1335, 47D1340, 47D1345, 47D1350, 47D1355, 47D1360, 47D1365, 47D1370, 47D1375, 47D1380, 47D1385, 47D1390, 47D1395, 47D1400, 47D1405, 47D1410, 47D1415, 47D1420, 47D1425, 47D1430, 47D1435, 47D1440, 47D1445, 47D1450, 47D1455, 47D1460, 47D1465, 47D1470, 47D1475, 47D1480, 47D1485, 47D1490, 47D1495, 47D1500, 47D1505, 47D1510, 47D1515, 47D1520, 47D1525, 47D1530, 47D1535, 47D1540, 47D1545, 47D1550, 47D1555, 47D1560, 47D1565, 47D1570, 47D1575, 47D1580, 47D1585, 47D1590, 47D1595, 47D1600, 47D1605, 47D1610, 47D1615, 47D1620, 47D1625, 47D1630, 47D1635, 47D1640, 47D1645, 47D1650, 47D1655, 47D1660, 47D1665, 47D1670, 47D1675, 47D1680, 47D1685, 47D1690, 47D1695, 47D1700, 47D1705, 47D1710, 47D1715, 47D1720, 47D1725, 47D1730, 47D1735, 47D1740, 47D1745, 47D1750, 47D1755, 47D1760, 47D1765, 47D1770, 47D1775, 47D1780, 47D1785, 47D1790, 47D1795, 47D1800, 47D1805, 47D1810, 47D1815, 47D1820, 47D1825, 47D1830, 47D1835, 47D1840, 47D1845, 47D1850, 47D1855, 47D1860, 47D1865, 47D1870, 47D1875, 47D1880, 47D1885, 47D1890, 47D1895, 47D1900, 47D1905, 47D1910, 47D1915, 47D1920, 47D1925, 47D1930, 47D1935, 47D1940, 47D1945, 47D1950, 47D1955, 47D1960, 47D1965, 47D1970, 47D1975, 47D1980, 47D1985, 47D1990, 47D1995, 47D2000, 47D2005, 47D2010, 47D2015, 47D2020, 47D2025, 47D2030, 47D2035, 47D2040, 47D2045, 47D2050, 47D2055, 47D2060, 47D2065, 47D2070, 47D2075, 47D2080, 47D2085, 47D2090, 47D2095, 47D2100, 47D2105, 47D2110, 47D2115, 47D2120, 47D2125, 47D2130, 47D2135, 47D2140, 47D2145, 47D2150, 47D2155, 47D2160, 47D2165, 47D2170, 47D2175, 47D2180, 47D2185, 47D2190, 47D2195, 47D2200, 47D2205, 47D2210, 47D2215, 47D2220, 47D2225, 47D2230, 47D2235, 47D2240, 47D2245, 47D2250, 47D2255, 47D2260, 47D2265, 47D2270, 47D2275, 47D2280, 47D2285, 47D2290, 47D2295, 47D2300, 47D2305, 47D2310, 47D2315, 47D2320, 47D2325, 47D2330, 47D2335, 47D2340, 47D2345, 47D2350, 47D2355, 47D2360, 47D2365, 47D2370, 47D2375, 47D2380, 47D2385, 47

FI MISC_Alliance 38

Misc difference 41

Misc difference 62

ET Misc Difference 63

ET Mic 21fforum 69

EST
M300 2144000000

PN - W09733911-A1.

PD 18-SEP-1997.
 PF 14-MAR-1997; U03461.
 PR 14-MAR-1996; US-615944.
 PA (UNIW) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI; 97-470818/43.

PT GDNF-neurturin family related growth factor, Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PS Claim 2; Page 150; 228pp; English.

PS Claim 2; Page 150; 228pp; English.

CC This polypeptide comprises mature mouse persephin, a novel growth
CC factor and member of the glial-derived neurotrophic factor-
CC neurturin family. Recombinant mature persephin can be expressed
CC in host cells utilising an isolated nucleic acid sequence (see
CC T90759). Persephin polypeptides, and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure. Persephin can be
CC added to a cell culture medium to promote growth and/or
CC differentiation.

| | | | | |
|-----------------------|--------|---------------------|--------|---------------|
| Query Match | 58.6%; | Score 584; | DB 26; | Length 89; |
| Best Local Similarity | 82.0%; | Pred. No. 2.43e-49; | | |
| Matches | 73; | Conservative | 12; | Mismatches 4; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

```

Db      1 crwsltipvaeilgyaseekvifrycagscpgeartqnsllvarllrgyrahypccg 60
        |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY     43 cqlmsltlsvalclgyaseekvifrycagscpgeartqhgltalrloggrahgppccr 102

```

Db 61 ptsyadvtflddqhmqqlpqlsaacgc 89
 |||:|||||:|||||
 QY 103 PTRTYDVAFLDDRRHRMQRLLPQLSAAACGC 131

| | |
|--------|----------------------------------|
| RESULT | 8 |
| ID | W30075 standard; Protein; 96 AA. |

DE Persephin-neurturin chimera (pan-growth factor).
 KW Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; haematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; mouse; PSP/NTN;
 KW pan-growth factor.

SECRET

CO. 4
UT 97073
77

Les Jours

label NIN

3300

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible][illegible]

11/11/2017 11:11 AM

DR WP1: 97-470818/43.
PT GDNF-neurturin family related growth factor, Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PS Claim 14; Page 174; 228pp; English.
CC This polypeptide comprises a chimera, named PSP/MTN, of mature
CC mouse persephin (see also W30066) and mature mouse neurturin.
CC It was obtained by PCR amplification (see T90808-11) of persephin
CC and neurturin sequences and expression in E. coli. Persephin and
CC neurturin are members of the glial-derived neurotrophic factor
CC family. The chimeric 'pan-growth factor' product is expected to
CC have the combined activities of persephin and neurturin. Pan-
CC growth factors (see also W30074), and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure.
SQ Sequence 96 AA;

| | | | | |
|-------------|-------------------|---------------------|--------|----------------|
| Query Match | 53.3%; | Score 531; | DB 26; | Length 96; |
| Best Local | Similarity 72.6%; | Pred. No. 1.05e-43; | | |
| Matches | 69; | Conservative | 13; | Mismatches 12; |
| | | | Indels | 1; |
| | | | Gaps | 1; |

```

Db      1 alagscrlwsltlpvaelglgyaseekvifrycagscpgeartqthslvlarlrgyrahg 60
        ||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY     38 ALSGPCQLMSLTLSVAELGLGYASEEKVIFRYCAGSCPRGARTQHTGLALARLOGGRAGH 97

```

```
db      61  rppcrptayedevsflvdhsryhtlgelsarecac  95
        ||||| | | : || | : | : | : || | | : |
QY      98  GPCCRPTRYTD-VAFLDDRHRMQRRLPQLSAACGC  131
```

| | |
|--------|-----------------------------------|
| RESULT | 9 |
| ID | W30076 standard; Protein; 100 AA. |

DT 27-MAR-1998 (first entry)
DE Neurturin-persephin chimera (pan-growth factor).
KW persephin; neurturin; glial-derived neurotrophic factor; GDNF;
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
KW Huntington's disease; nervous system tumour; multiple sclerosis;
KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
KW congestive heart failure; therapy; mouse; NTN/psp;
KW pan-growth factor.

| OS | Chimeric - Mus musculus. | location/Qualifiers |
|----|--------------------------|---------------------|
| FH | key | |
| FT | protein | 1..67 |

| | | | |
|----|---------|---|--|
| FT | /label- | NTN | |
| FT | /note- | "residues 1-67 of mature murine neuriturin" | |
| FT | 68..100 | | |
| FT | /label- | PSP | |
| FT | /note- | "residues 64-96 of mature murine persephin" | |

PN WO9733911-A1.
PD 18-SEP-1997.
PE 14-MAR-1997; U03461.
PR 14-MAR-1996; US-615944.
PA (UNIW) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI, 97-470818/43.
PT GDNF-neurturin family related growth factor, Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PS Claim 14; Page 175-176; 228pp; English.

CC This polypeptide comprises a chimera, named NTN/PSP, of mature
CC mouse neurturin and mature mouse persephin (see W30066). It was
CC obtained by PCR amplification (see T90808, T90811-13) of persephin
CC and neurturin sequences and expression in *E. coli*. Persephin and
CC neurturin are members of the glial-derived neurotrophic factor
CC family. The chimeric 'pan-growth factor' product is expected to
CC have the combined activities of persephin and neurturin. Pan-
CC growth factors (see also W30073), and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure.

| | | | | |
|-----------------------|--------|---------------------|--------|----------------|
| Query Match | 41.6%; | Score 415; | DB 26; | Length 100; |
| Best Local Similarity | 58.3%; | Pred. No. 1.77e-31; | | |
| Matches | 56; | Conservative | 19; | Mismatches 17; |
| | | | | Indels 4; |
| | | | | Gaps 4; |

```
Db      5 pcglrelevrwsejlgysdetsvlrfycagaceaalrlydlgrlrqrrvrrerara 64
        ||| : | : ||||| : | : ||||| : | : | : | : | : | : ||::
QY     42 PCQLMSLTSLVAELGLGYASEEKVIFRYCAGSCPRGART-OHGIA-L-ARLOGO-GAAGH 97
```

```
D6  hpcqptsyadvrtfiddqghwqgipqlsaacycg 100
    |||::|||::||||:::|||::|
QY  98 GPCCRPTFRYTDAFLDDRHRKWRQLPOLSAACGCG 133
```

| | |
|--------|----------------------------------|
| RESULT | 10 |
| ID | W13717 standard; Protein; 195 AA |

DT 09-FEB-1998 (first entry)
DE Mouse pre-pro-neurturin.
KW Neurturin; mouse; haematopoietic cell; neuronal cell; stem cell; NT gene; neurodegenerative disease; peripheral neuropathy; nervous system tumour;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; ischaemic stroke; acute brain injury; basopaenia;
KW acute spinal cord injury; multiple sclerosis; eosinopaenia; lymphopaenia;
KW monocytopenia; neutropaenia; anaemia; thrombocytopaenia; neuroblastoma;
KW antibody; obesity; therapy.
OS Mus musculus.

| Key | Location/Qualifiers |
|------------|--------------------------|
| FT Region | 1..19 |
| FT Region | /note= "pre-region" |
| FT Region | 20..95 |
| FT Region | /note= "pro-region" |
| FT Peptide | 1..95 |
| FT Peptide | /note= "pre-pro-peptide" |

PN MO9708196-A1.
PD 06-MAR-1997. U14065.
PF 27-AUG-1996; US-519777.
PR 28-AUG-1995; US-519777.
PA (UNIW) UNIV WASHINGTON.
PI Johnson EM, Koltbauer PT, Lampe PA, Milbrandt JD,
DR WPI; 97-179176/16.

PT A novel growth factor Neurturin - used to treat neuro-degenerative
PT and haematopoietic cell degeneration diseases, e.g. Alzheimer's
PT disease and eosinopenia
PS Example 6; Page 108; 206pp; English.
CC This sequence represents the pre-pro form of mouse neurturin (NT). NT
CC promotes the growth and differentiation of haematopoietic and neuronal
CC cells, and their stem cells. The NT gene and protein are used to prevent
CC or treat neurodegenerative diseases e.g. peripheral neuropathy,
CC amyotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease,
CC Huntington's disease, ischaemic stroke, acute brain injury, acute spinal


```

RESULT      13
ID    W13710 standard; Protein; 102 AA.
AC     W13710;
DT     09-FEB-1998   (first entry)
DE     Mature human neuritin.
KW     Neuritin; human; haematopoietic cell; neuronal cell; stem cell; NT gene; neurodegenerative disease; peripheral neuropathy; nervous system tumour; amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease; Huntingdon's disease; ischemic stroke; acute brain injury; baso-paenia; acute spinal cord injury; multiple sclerosis; eosinopaenia; lymphopae-nia; monocytopenia; neutropenia; anaemia; thrombocytopaenia; neuroblastoma; antibody; obesity; therapy.
KM     Homo sapiens.
SW
FH     Key          Location/Qualifiers
FT           Region       8..101
FT                                     /note= "specifically claimed minimum active fragment"
PN         WO9708196-A1.
PD         06-MAR-1997.
PF         27-AUG-1996; U14065.
PR         28-AUG-1995; US-519777.
PA         (UNIW ) UNIV WASHINGTON.
PI         Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR         MPI; 97-179176/16..
NR         N-PSDB; T61468.
PT         A novel growth factor Neuritin - used to treat neuro-degenerative and haematopoietic cell degeneration diseases, e.g. Alzheimer's disease and eosinopenia
PT         Claim 3; Page 104-105; 206pp; English.
PC         This sequence represents the mature form of human neuritin (NT). NT promotes the growth and differentiation of haematopoietic and neuronal cells, and their stem cells. The NT gene and protein are used to prevent or treat neurodegenerative diseases e.g. peripheral neuropathY, amytrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's disease, ischemic stroke, acute brain injury, acute spinal cord injury, nervous system tumors, multiple sclerosis and infection; CC and haematopoietic cell degenerative diseases, e.g. eosinopaenia, basopaenia, lymphopoenia, monocytopenia, neutropaenia, anaemia, thrombocytopenia and stem cell insufficiencies. The NT protein and gene are also useful to treat neuroblastomas. Antibodies against NT and oligonucleotides (used as either probes or primers, corresponding to an exon of pre-pro-NT gene or flanking a target sequence) can be used for detecting NT in a sample or detecting mutations in the NT gene. Antisense sequences of the NT gene are used to treat diseases promoted by NT expression e.g. obesity.
CC         Sequence 102 AA;
SQ
Query Match             32.6%; Score 325; DB 25; Length 102;
Best Local Similarity 49.5%; Pred. No. 3.80e-22;
Matches 47; Conservative 18; Mismatches 25; Indels 5; Gaps 5;

Db        7 pcdlrelelvseglgyasdetvlfrycagaceaaarvydlgrlrqrllrreryva 66
||| : | : ||||||| : | : ||||| : | : | : | : | : | : | :
Oy        42 PCQLWSLTLSVAELGLGYASEKVIFFRYCAGSCPRGART-QHGILA-L-ARLO-GQGRAHC 97

Db        67 qpccrptayedevsfidahsrhythvtshlsarecac 101
||||| | | : ||| : | : : ||| | : |
Oy        98 GPCCRPTRTYT-D-VAFLLDRLHRMQRLLPOLSAACGC 131


RESULT      14
ID    W15744 standard; Protein; 104 AA.
AC     W15744;
DT     28-NOV-1997   (first entry)
DE     (Arg32-Ile134) truncated GDNF.
KW     Glial cell line-derived neurotrophlc factor; GDNF; human; dopaminergic; nerve cell; Parkinson's disease; gene therapy.
OS     Chimeric Homo sapIiens;
OS     Chimeric synthetic.
PN         WO9711964-A1.
PF         03-APR-1997.
PR         16-SEP-1996; U14915.
PR         28-SEP-1995; US-535681.
```

```
PA (AMGE-) AMGEN INC.  
PI Hu SS;  
DR WPI; 97-212849/19.  
DR N-PSTDB; T60545.  
PT Truncated glial cell line-derived neurotrophic factor protein - used  
PT In the treatment and gene therapy of Parkinson's disease  
PS Example 3; Page 77; 105pp; English.  
CC This polypeptide comprises (Arg32-Ile134)-truncated glial cell  
CC line-derived neurotrophic factor (GDNF) protein. It can be  
CC expressed in host cells using a vector carrying a truncated GDNF  
CC polynucleotide (see T60545) produced by recombinant genetic  
CC engineering techniques. Claimed truncated GDNFs can be used in the  
CC treatment of nervous system damage caused by disease or injury, esp.  
CC in the treatment of Parkinson's disease. Also claimed are the  
CC vector, the host cell (pref. E. coli or CHO), the polynucleotide  
CC and a composition comprising mature GDNF (see W15706) and one or  
CC more truncated GDNFs of mol.wt. 29-40 kDa. The polynucleotide can  
CC be used to provide in vivo production of truncated GDNF, thereby  
CC providing gene therapy.  
SQ Sequence 104 AA;
```

Query Match 24.5%; Score 244; DB 24; Length 104;
Best Local Similarity 39.2%; Pred. No. 6.08e-14;
Matches 40; Conservative 20; Mismatches 38; Indels 4; Gaps 4

Dd 2 rggrgknrgcvltahlnvdtlglgetkeellfrycsgscdaettydkllnlsmnr 61
I::: : I | ::|::| |||| : | : |||||::| : | : | :
QY 34 RLRRALSGPCQLMSLTLSVAELGLGYASEKVIIFRYCAGSCPRGART-QHGL-ALARLQG 91
.:.||| : | ::|||| : | || |||

Dd 62 lvsdvkgagacrpiafddlsfidnlyhlrlkhsakrcgc 103
|.:||| : | ::|||| : | || |||

QY 92 QGRAHG-G-PCCRPTYTD-VAFLDRLRWRQRLLPQLSAAACGC 131

RESULT 15
ID W15743 standard; Protein; 114 AA.

AC W15743;
DT 28-NOV-1997 (first entry)
DE (Pro23-Lys37delasn37-Ile-134) truncated GDNF.
KW Glial cell line-derived neurotrophic factor; GDNF; human;
KM dopaminergic; nerve cell; Parkinson's disease; gene therapy.
OS Chimeric Homo sapiens;
SS Chimeric synthetic.
PN WO9711964-A1.
PD 03-APR-1997.
PF 16-SEP-1996; U14915.
PR 28-SEP-1995; US-535681.
PA (AMGE-) AMGEN INC.
PI Hu SS;
DR WPI; 97-212849/19.
DR N-PSTDB; T60544.
PT Truncated glial cell line-derived neurotrophic factor protein - used
PT In the treatment and gene therapy of Parkinson's disease
PS Example 3; Page 75-76; 105pp; English.
CC This polypeptide comprises (Pro23-Lys37delasn37-Ile134)-truncated
CC glial cell line-derived neurotrophic factor (GDNF) protein. It
CC can be expressed in host cells using a vector carrying a truncated
CC GDNF polynucleotide (see T60544) produced by recombinant genetic
CC engineering techniques. Claimed truncated GDNFs can be used in the
CC treatment of nervous system damage caused by disease or injury, esp.
CC in the treatment of Parkinson's disease. Also claimed are the
CC vector, the host cell (pref. E. coli or CHO), the polynucleotide
CC and a composition comprising mature GDNF (see W15706) and one or
CC more truncated GDNFs of mol.wt. 29-40 kDa. The polynucleotide can
CC be used to provide in vivo production of truncated GDNF, thereby
CC providing gene therapy.
SQ Sequence 114 AA;

Query Match 24.5%; Score 244; DB 24; Length 114;
Best Local Similarity 39.2%; Pred. No. 6.08e-14;
Matches 40; Conservative 20; Mismatches 38; Indels 4; Gaps 4

Dd 12 rggrrgnrvcvtaihlnvdtlglgetkeellfrycsgscdaettydkllnlsmnr 71

Wed Sep 29 14:58:50 1999

US-08-931-858E-132.rag

Page 8

[illegible]

Search completed: Wed Sep 29 14:01:32 1999
Job time : 31 secs.

10

OY 76 RGAR-TQHGALARLQGGRAGHG-PCCRPTRYTD-VAF-LDDRHRM-QRLPQLSAACG 130
Db 552 C 552
OY 131 C 131

RESULT 5
ENTRY #type complete
TITLE Mullerian inhibiting factor precursor - bovine
ALTERNATE_NAMES Mullerian inhibiting substance (MIS)
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 12-Apr-1996

ACCESSIONS A01398; B01398
REFERENCE A90879
#authors Cate, R.L.; Mattaliano, R.J.; Hession, C.; Tizard, R.;
Farber, N.M.; Cheung, A.; Ninfa, E.G.; Frey, A.Z.; Gash,
D.J.; Chow, E.P.; Fisher, R.A.; Bertoni, J.M.; Torres, G.;
Wallner, B.P.; Ramachandran, K.L.; Ragin, R.C.; Mangano, T.F.; MacLaughlin, D.T.; Donahoe, P.K.

#journal Cell (1986) 45:685-698
#title Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells.

#cross-references MUID:86218082
#accession A01398
#molecule_type DNA
#residues 1-14 #label CAL
#experimental_source newborn calf testis, clones cbm15 and ps21
#accession B01398
#molecule_type mRNA

COMMENT ##residues 15-575 #label CA2
This glycoprotein, produced by the Sertoli cells of the testis, causes regression of the Mullerian duct. It also is able, in vivo and in vitro, to inhibit the growth of tumors derived from tissues of Mullerian duct origin. Other roles for this protein in gonadal differentiation, meiosis inhibition, and testicle descent are suggested by the low MIS concentrations found in the testis after duct regression and in the adult ovary.

COMMENT This protein is homologous to the beta transforming growth factor, inhibin alpha chain, and inhibin beta A and B chains. The best area of homology corresponds to the mature protein of all these sequences. All of these proteins are biologically active as disulfide-linked dimers.

COMMENT Although it does not compete with EGF for receptor binding sites, MIS can inhibit the autophosphorylation of the EGF receptor in vitro.

CLASSIFICATION #superfamily inhibin
KEYWORDS #superfamily inhibin
FEATURES cytoxin; glycoprotein; gonadal differentiation; testis

1-19 #domain signal sequence #status predicted #label SIG\
20-24 #domain propeptide #status predicted #label PRO\
25-575 #product Mullerian inhibiting factor #status predicted
#label MAT\
78,344 #binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 575 #molecular-weight 60623 #checksum 8089
Query Match 11.5%; Score 115; DB 1; Length 575;
Best Local Similarity 32.8%; Pred. No. 3.85e-05;
Matches 20; Conservative 17; Mismatches 19; Indels 5; Gaps 5;

Db 514 RNPRTGNHVVLLKMQARGATLAPPCVPTAYTGKLLISERISAHHPNVATECG 573
OY 76 RGAR-TQHGALARLQGGRAGHG-PCCRPTRYTD-VAF-L-DDRHRMQLPQLSAACG 130

Db 574 C 574
OY 131 C 131

RESULT 6
ENTRY #type complete
TITLE Mullerian inhibiting factor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Sep-1997

ACCESSIONS S20100; S51159
REFERENCE S20100

#authors Muensterberg, A.; Lovell-Badge, R.
#journal Development (1991) 113:613-624
#title Expression of the mouse anti-Mullerian hormone gene suggests a role in both male and female sexual differentiation.

#cross-references MUID:92146272
#accession S20100
#molecule_type DNA
#residues 1-555 #label MUE
#cross-references EMBL:X63240; NID:g49945; PID:g49946
REFERENCE S51159

#authors Dresser, D.W.; Hacker, A.; Lovell-Badge, R.; Guerrier, D.
#submission Submitted to the EMBL Data Library, January 1995
#description The genes for anti-Mullerian hormone (AMH) and a spliceosome protein (SAP62) are contiguous.

#accession S51159
#status Preliminary
#molecule_type DNA
#residues 1-41 #label DRE
#cross-references EMBL:X83733

GENETICS #introns 135/1; 182/3; 219/1; 272/2
CLASSIFICATION #superfamily inhibin
SUMMARY #length 555 #molecular-weight 59778 #checksum 3223

Query Match 11.4%; Score 114; DB 2; Length 555;
Best Local Similarity 34.4%; Pred. No. 5.69e-05;
Matches 21; Conservative 16; Mismatches 19; Indels 5; Gaps 5;

Db 494 RNPRTGNHVVLLKMQARGALGRLPCCVPTAYAGKLLISERISADHVPNVATECG 553
OY 76 RGAR-TQHGALARLQGGRAGHG-PCCRPTRYTD-VAF-L-DDRHRMQLPQLSAACG 130

Db 554 C 554
OY 131 C 131

RESULT 7
ENTRY #type complete
TITLE Mullerian inhibiting factor precursor - human
ALTERNATE_NAMES anti-Mullerian hormone; mullerian inhibiting substance (MIS)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 29-May-1998

ACCESSIONS A01397
REFERENCE A90879
#authors Cate, R.L.; Mattaliano, R.J.; Hession, C.; Tizard, R.;
Farber, N.M.; Cheung, A.; Ninfa, E.G.; Frey, A.Z.; Gash,
D.J.; Chow, E.P.; Fisher, R.A.; Bertoni, J.M.; Torres, G.;
Wallner, B.P.; Ramachandran, K.L.; Ragin, R.C.; Mangano, T.F.; MacLaughlin, D.T.; Donahoe, P.K.

#journal Cell (1986) 45:685-698
#title Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells.

#cross-references MUID:86218082
#accession A01397
#molecule_type DNA
#residues 1-560 #label CAT

COMMENT ##cross-references GB:K03474; NID:q188560; PID:g386953
Although it does not compete with EGF for receptor binding sites, MIS can inhibit the autophosphorylation of the EGF receptor in vitro.
COMMENT For anti-Mullerian hormone type II receptor, see PIR:JC4335.
GENETICS

[illegible]

```

#title      Complementary deoxyribonucleic acid/cloning of a novel
            transforming growth factor-beta messenger ribonucleic acid
            from chick embryo chondrocytes.
#cross-references MUID:89096966
#accession.    A34939
##status      preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues    1-412 ##label JAK
##cross-references GB:M31154; NID:g212758; PID:g212759
S25850
REFERENCE
#authors      Burt, D.W.; Paton, I.R.; Dey, B.R.
#journal      J. Mol. Endocrinol. (1991) 7:175-183
#title        Comparative analysis of human and chicken transforming growth
            factor-beta-2 and -beta-3 promoters.
#cross-references MUID:92134496
#accession    S25850
##status      preliminary
##molecule_type DNA
##residues    1-117 ##label BUR
##cross-references EMBL:X58127; NID:g63815; PID:g63816
#accession    S36125
##status      preliminary; nucleic acid sequence not shown;
            translation not shown
##molecule_type DNA
##residues    119-172 ##label BU2
##cross-references EMBL:X60055; NID:g396688; PID:g396689
##note        the nucleotide sequence was submitted to the EMBL Data
            Library, June 1991
#accession    S36124
##status      preliminary; nucleic acid sequence not shown;
            translation not shown
##molecule_type DNA
##residues    173-322, 'ELPT', 327-412 ##label BU3
##cross-references EMBL:X60091
##note        the nucleotide sequence was submitted to the EMBL Data
            Library, June 1991
REFERENCE
#authors      I51181
            Jakowlew, S.B.; Lechleider, R.; Geiser, A.G.; Kim, S.J.;
            Santa-Coloma, T.A.; Cubert, J.; Sporn, M.B.; Roberts, A.B.
#journal      Mol. Endocrinol. (1992) 6:1285-1298
#title        Identification and characterization of the chicken
            transforming growth factor-beta 3 promoter.
#cross-references MUID:93024487
#accession    I51181
##status      preliminary
##molecule_type DNA
##residues    1-117 ##label JA2
##cross-references GB:S46000; NID:g257172; PID:g257173
GENETICS
#introns      216/1; 252/1; 309/2; 360/3
#note         list of introns may be incomplete
CLASSIFICATION
#superfamily inhibin
KEYWORDS      glycoprotein; growth factor; growth regulation; homodimer;
            mitogen
FEATURE
1-22         #domain signal sequence #status predicted #label SIG\
23-300      #domain propeptide #status predicted #label PRO\
301-412     #product transforming growth factor beta-3 #status
            predicted #label MAR\
74, 142     #binding_site carbohydrate (Asn) (covalent) #status
            predicted
SUMMARY      #length 412 #molecular-weight 47183 #checksum 2660
Query Match 10.3%; Score 103; DB 2; Length 412;
Best Local Similarity 29.8%; Pred. No. 3.68e-03;
Matches 28; Conservative 18; Mismatches 43; Indels 5; Gaps 4;
Db 309 RNLEBNCVRLPYIDFRQDLCGKWKVHEPKGYFANFCGSPCPYLRASADTHTSTVLGLVNTL 368
QY 37 RALSGPCQLWSLTLSV-AELGLGYASEKVIIF-RYCAGSCP--RGARTQHGIALARLQGG 92
Db 369 NPEASASPCCVQDLEPLTILYVGRTPKVEQLS 402

```


QY 93 GR-AHGGPCCRPRTYTDVAFLLDRHRWQRLPQLS 125

RESULT 10

ENTRY D46352 #type complete

TITLE ORF4 protein - Chlorella virus PBCV-1

ORGANISM #formal_name Chlorella virus PBCV-1

DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

ACCESSIONS D46352

REFERENCE A46352

#authors Schuster, A.M.; Graves, M.; Korth, R.; Ziegelbein, M.; Brumbaugh, J.; Grone, D.; Meints, R.H.

#journal Virology (1990) 176:515-523

#title Transcription and sequence studies of a 4.3-kbp fragment from a ds-DNA eukaryotic algal virus.

#cross-references MUID:90266467

#accession D46352

#status preliminary

#molecule_type DNA

#residues 1-265 #label SCH

##cross-references GB:M33758; NID:g323370; PID:g807596

SUMMARY #length 265 #molecular-weight 31237 #checksum 1642

Query Match 10.0%; Score 100; DB 2; Length 265;

Best Local Similarity 28.8%; Pred. No. 1.10e-02;

Matches 15; Conservative 18; Mismatches 15; Indels 4; Gaps 4;

Db 180 THKISLGHKRTMSTICPFMWSLTSTTVKKNVLFNAEEYNLYRI-SQDCPK 230

QY 28 THR-PLARLRALSGPCQ-LWSLT-LSVAELGLGYASEEKVIFRYCAGSCPR 76

RESULT 11

ENTRY A41397 #type complete

TITLE transforming growth factor beta-3 precursor - mouse

ORGANISM #formal_name Mus musculus #common_name house mouse

DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 17-Mar-1999

ACCESSIONS A41397; A61039; A61225

REFERENCE A41397

#authors Miller, D.A.; Lee, A.; Matsui, Y.; Chen, E.Y.; Moses, H.L.; Derynck, R.

#journal Mol. Endocrinol. (1989) 3:1926-1934

#title Complementary DNA cloning of the murine transforming growth factor-beta3 (TGFbeta3) precursor and the comparative expression of TGFbeta3 and TGFbeta1 messenger RNA in murine embryos and adult tissues.

#cross-references MUID:90190650

#accession A41397

#molecule_type mRNA

#residues 1-410 #label MIL

##cross-references GB:M32745; NID:g201949; PID:g201950

REFERENCE A61039

#authors Denhez, F.; Lafyatis, R.; Kondaliah, P.; Roberts, A.B.; Sporn, M.B.

#journal Growth Factors (1990) 3:139-146

#title Cloning by polymerase chain reaction of a new mouse TGF-beta, mTGF-beta3.

#cross-references MUID:91000714

#accession A61039

#molecule_type mRNA

#residues 1-410 #label DEN

REFERENCE A61225

#authors Martin, F.; Scotto, L.; Assoian, R.K.; Wolgemuth, D.J.

#journal Cell Growth Differ. (1991) 2:77-83

#title Cell lineage specificity of expression of the murine transforming growth factor beta-3 and transforming growth factor beta-1 genes.

#cross-references MUID:91299576

#accession A61225

#status translation not shown

#molecule_type mRNA

##residues 285-410 #label WAT

CLASSIFICATION #superfamily inhibin

KEYWORDS glycoprotein; growth factor; growth regulation

FEATURE 1-21

22-298

259-261

299-410

72,133,140

#domain signal sequence #status predicted #label SIG\

#domain propeptide #status predicted #label PRO\

#region cell attachment (R-G-D) motif\

#product transforming growth factor beta-3 #status predicted #label MAT\

#binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 410 #molecular-weight 46884 #checksum 8579

Query Match 9.9%; Score 99; DB 2; Length 410;

Best Local Similarity 28.7%; Pred. No. 1.58e-02;

Matches 27; Conservative 19; Mismatches 43; Indels 5; Gaps 4;

Db 307 RNLENCVRLPYIDFRODLGKWKVHEPKGYANPCSGPCPYLRSDYTHSTVGLYNTL 366

QY 37 RALSGPCQLMSLTLSV-AELGLGYASEEKVIF-RYCAGSCP--RGARTQHGIALARLQGG 92

Db 367 NPEASAPCCVPQDLEPLTILYVGRTPKVEQLS 400

QY 93 GR-AHGGPCCRPRTYTDVAFLLDRHRWQRLPQLS 125

RESULT 12

ENTRY A36169 #type complete

TITLE transforming growth factor beta-3 precursor - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 14-Dec-1990 #sequence_revision 14-Dec-1990 #text_change 24-Sep-1998

ACCESSIONS A36169; A41262; S01824

REFERENCE A36169

#authors ten Dijke, P.; Hansen, P.; Iwata, K.K.; Pieler, C.; Foulkes, J.G.

#journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:4715-4719

#title Identification of another member of the transforming growth factor type beta gene family.

#cross-references MUID:88263019

#accession A36169

#status preliminary; not compared with conceptual translation

#molecule_type mRNA

#residues 1-412 #label TEN

##cross-references GB:J03241; NID:g339551; PID:g339552

REFERENCE A41262

#authors Arrick, B.A.; Lee, A.L.; Grendell, R.L.; Derynck, R.

#journal Mol. Cell. Biol. (1991) 11:4306-4313

#title Inhibition of translation of transforming growth factor-beta3 mRNA by its 5' untranslated region.

#cross-references MUID:91342629

#accession A41262

#molecule_type DNA

#residues 1-48 #label ARR

##cross-references GB:M58524

REFERENCE S01824

#authors Derynck, R.; Lindquist, P.B.; Lee, A.; Wen, D.; Tamm, J.; Graycar, J.L.; Rhee, L.; Mason, A.J.; Miller, D.A.; Coffey, R.J.; Moses, H.L.; Chen, E.Y.

#journal EMBO J. (1988) 7:3737-3743

#title A new type of transforming growth factor-beta, TGF-beta3.

#cross-references MUID:89091120

#accession S01824

#molecule_type mRNA

#residues 3-412 #label DER

##cross-references EMBL:X14885; NID:g37075; PID:e223372; PID:g1200236

GENETICS

#gene GDB:TGFB3

#cross-references GDB:120437; OMIM:190230

#map_position 14q24-14q24

#introns 118/1; 172/3; 216/1; 252/1; 309/2; 360/3

CLASSIFICATION #superfamily inhibin

KEYWORDS growth factor; homodimer


```
#domain signal sequence #status predicted #label sig\
#domain propeptide #status predicted #label PRO\
#product transforming growth factor beta-3 #status
predicted #label MAT
SUMMARY          #length 412 #molecular-weight 47328 #checksum 2862

Query Match      9.9%; Score 99; DB 2; Length 412;
Best Local Similarity 28.7%; Pred. No. 1.58e-02;
Matches 27; Conservative 19; Mismatches 43; Indels 5; Gaps 4;

Db   309 RNLENCVPRPLIDFRQDLGKWKWHEPKGIYANFGSGPCPYLRADTTHSTVLGLYNTL 368
    | | :|::||: || | | : |:|:|| | | | | :|
Oy   37 RALSGPQLWSLTSLV-AELGLGYASEKEVI-F-RYCAGSCP--RGARTQHGLARLGQG 92
                                     ::|| | |::|||

Db   369 NPEASAPCCVPQDLEPLETLYYVGRTPKVEQLS 402
    |::||| |::|| |::|||
Oy   93 GR-AHGPPCRRPTRYTDVAFLDDRHRWRQLPOLIS 125

RESULT 13
ENTRY   #type complete
TITLE   probable lyss protein - Mycobacterium tuberculosis (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE     17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS
REFERENCE A70954
         A70500
AUTHORS   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekaita, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
        Nature (1998) 393:537-544
        Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
        #cross-references MUID:98295987
        #accession G70954
        #status preliminary; nucleic acid sequence not shown; translation not shown
        ##molecule_type DNA
        ##residues 1-505 ##label COL
        ##cross-references GB:Z95557; GB:AL123456; NID:g3242276; PID:e316848; PID:g2113978
        ##experimental_source strain H37Rv
GENETICS
#gene lyss
SUMMARY #length 505 #molecular-weight 55709 #checksum 8894

Query Match      9.9%; Score 99; DB 2; Length 505;
Best Local Similarity 35.9%; Pred. NO. 1.58e-02;
Matches 14; Conservative 10; Mismatches 14; Indels 1; Gaps 1;

Db   399 PLTRQHSIPGVTEKNLDYLRLGIELATGYSELSDPVVQR 437
    ||:|:|:|:| | | | | | | | | | | | | | |
Oy   31 PLARLRALSGPQLMSLTLSVAELGLGYAS-EKKIVIR 68

RESULT 14
ENTRY   #type complete
TITLE   transforming growth factor beta-3 precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE     03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change
ACCESSIONS A55706; B40699; S36042
REFERENCE A55706
AUTHORS   Wang, J.; Kuliszewski, M.; Yee, W.; Sedlackova, L.; Xu, J.;
```

```
#journal      J. Biol. Chem. (1995) 270:2722-2728  
#title       Cloning and expression of glucocorticoid-induced genes in fetal rat lung fibroblasts. Transforming growth factor-beta-3.  
  
#cross-references MUID:95155340  
#accession     A55706  
##molecule_type mRNA  
##residues    1-410 ##label WAN  
##cross-references GB:U03491  
##note        it is uncertain whether Met-1 is the initiator  
  
REFERENCE  
A40699  
McKinnon, R.D.; Piras, G.; Ida Jr., J.A.; Dubois-Dalcq, M.  
J. Cell Biol. (1993) 121:1397-1407  
A role for TGF-beta in oligodendrocyte differentiation.  
B40699  
Preliminary  
##status      preliminary  
##molecule_type mRNA  
##residues    157-211 ##label MCK  
##cross-references EMBL:X71903; NID:g311326; PID:g311327  
CLASSIFICATION #superfamily inhibitin  
SUMMARY length 410 #molecular-weight 46856 #checksum 8897
```

Query Match 9.7%; Score 97; DB 2; Length 410;
Best Local Similarity 27.7%; Pred.No. 3.24e-02;
Matches 26; Conservative 20; Mismatches 43; Indels 5; Gaps 4;

Dn 307 RNIEENCCVRPLYIDFRODIGNKWVHEPKGYANFCGCPYLRSSDTTHSTVLGLYNLT 366
 | | :|::|| : || | : |:|:|| | :: | |:
Qy 37 RAISGPCQLMSLTLSV-AELGIAYASEKVIF-RYCAGSCP--RGARTQHGLALARLGQ 92
 |::||| | :|::| |::| |
Db 367 NPEASAPPCCVPDLEPTLIYYVGRTPKVEOLS 400
 |::||| | :|::| |::| |
Qy 93 GR-ANGGCCRPTRTYTDVAFLDDRHRWQLPOLSL 125

RESULT 15
ENTRY S01825 #type complete
TITLE transforming growth factor beta-3 precursor - pig
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 08-Sep-1997

ACCESSIONS
REFERENCE S01825
#authors Derynck, R.; Lindquist, P.B.; Lee, A.; Wen, D.; Tammm, J.; Graycar, J.L.; Rhoe, L.; Mason, A.J.; Miller, D.A.; Coffey, R.J.; Moses, H.L.; Chen, E.Y.

#journal EMBO J. (1988) 7:3737-3743
#title A new type of transforming growth factor-beta, TGF-beta3.
#cross-references MUID:89091120
#accession S01825
##molecule_type mRNA
##residues 1-409 ##label DER
##cross-references EMBL:X14150; NID:g2127; PID:g2128
CLASSIFICATION #superfamily inhibin
KEYWORDS growth factor

FEATURE
1-25 #domain signal sequence #status predicted #label SIG\
26-297 #domain propeptide #status predicted #label PRO\
298-409 #product transforming growth factor beta-3 #status predicted #label MAT

SUMMARY length 409 #molecular-weight 46814 #checksum 7701

Query Match 9.6%; Score 96; DB 2; Length 409;
Best Local Similarity 30.6%; Pred.No. 4.62e-02;
Matches 22; Conservative 14; Mismatches 31; Indels 5; Gaps 4;

Dn 306 RNIEENCCVRPLYIDFRODIGMKRWVHEPKGYANFCGCPYLRSADTTHTSVLGLYNLT 365
 | | :|::|| : || | : |:|:|| | :: | |:
Qy 37 RAISGPCQLMSLTLSV-AELGIAYASEKVIF-RYCAGSCP--RGARTQHGLALARLGQ 92
 |::||| | :|::| |::| |
Db 366 NPEASAPPCVP 377

Wed Sep 29 14:58:51 1999

US-08-931-858E-132.rpt

Page 7

OY 93 GR-AHGGPCCR 103

Search completed: Wed Sep 29 14:00:44 1999
Job time : 14 secs.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:14 ; Search time 1811.29 Seconds
(without alignments)
1588.882 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTGAGCCGCACACTGTTC.....TCAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database : EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| C 1 | 358.4 | 24.6 | 360 | 50 | AI675523 | AI675523 wc01f01.x |
| 2 | 286.4 | 19.6 | 777 | 48 | AI557281 | AI557281 PT2.1_15_ |
| C 3 | 245.4 | 16.8 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 4 | 244.6 | 16.8 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 5 | 242.8 | 16.6 | 722 | 48 | AI557025 | AI557025 PT2.1_10_ |
| C 6 | 226.2 | 15.5 | 481 | 50 | AI669421 | AI669421 ty32d03.x |
| 7 | 225.4 | 15.4 | 415 | 48 | AI609972 | AI609972 tt78c10.x |
| 8 | 223 | 15.3 | 407 | 35 | AA535216 | AA535216 nj75f05.s |
| C 9 | 222.4 | 15.2 | 479 | 26 | W96522 | W96522 ze43f08.r1 |
| C 10 | 222.2 | 15.2 | 437 | 36 | AA644090 | AA644090 ab62b04.s |
| C 11 | 220.2 | 15.1 | 500 | 36 | AA626040 | AA626040 af94a08.s |
| C 12 | 220.2 | 15.1 | 454 | 49 | AI634187 | AI634187 ts55a08.x |
| C 13 | 220 | 15.1 | 374 | 38 | AA746911 | AA746911 nx69c10.s |
| 14 | 219.8 | 15.1 | 434 | 36 | AA630854 | AA630854 nt57f09.s |
| C 15 | 219.6 | 15.1 | 388 | 39 | AA837686 | AA837686 oe05h09.s |
| C 16 | 219.4 | 15.0 | 323 | 35 | AA536040 | AA536040 nj81c09.s |
| C 17 | 219.2 | 15.0 | 484 | 33 | AA410788 | AA410788 zt35b11.r |
| 18 | 219.2 | 15.0 | 380 | 36 | AA613624 | AA613624 no18d03.s |
| C 19 | 218.4 | 15.0 | 458 | 51 | AI733856 | AI733856 zo19c03.y |
| C 20 | 218.2 | 15.0 | 408 | 34 | AA486877 | AA486877 ab16h04.s |
| C 21 | 218 | 14.9 | 533 | 39 | AA833875 | AA833875 od64e08.s |
| 22 | 217.2 | 14.9 | 470 | 34 | AA456924 | AA456924 aa90b09.s |
| C 23 | 217.2 | 14.9 | 329 | 36 | AA643770 | AA643770 np06e10.s |
| C 24 | 217.2 | 14.9 | 466 | 42 | AI087040 | AI087040 oy70a10.x |
| C 25 | 217.2 | 14.9 | 365 | 46 | AI419337 | AI419337 tf27h01.x |
| C 26 | 217.2 | 14.9 | 397 | 46 | AI421950 | AI421950 tf45d05.x |
| 27 | 217 | 14.9 | 415 | 34 | AA515048 | AA515048 ng67h10.s |
| C 28 | 216.6 | 14.8 | 461 | 30 | AA225406 | AA225406 nc24d02.r |
| 29 | 216.6 | 14.8 | 301 | 39 | AA828592 | AA828592 od74e10.s |
| C 30 | 216.4 | 14.8 | 528 | 39 | AA833896 | AA833896 od64g08.s |
| C 31 | 216.4 | 14.8 | 344 | 41 | AI054030 | AI054030 qi66a12.x |
| C 32 | 216.2 | 14.8 | 428 | 41 | AI066646 | AI066646 oz82d10.x |
| C 33 | 216 | 14.8 | 569 | 53 | HSM007223 | AI042373 Homo sapi |
| C 34 | 215.6 | 14.8 | 367 | 45 | AI361090 | AI361090 gy04d06.x |
| C 35 | 215.6 | 14.8 | 524 | 47 | AI524240 | AI524240 th11d12.x |
| C 36 | 215.4 | 14.8 | 457 | 31 | AA284247 | AA284247 zc65d01.T |
| C 37 | 215.4 | 14.8 | 480 | 49 | AI635028 | AI635028 tz03d06.x |
| C 38 | 215.2 | 14.7 | 376 | 29 | AA176978 | AA176978 nc01c01.s |
| C 39 | 215.2 | 14.7 | 474 | 46 | AI457313 | AI457313 tl173a05.x |
| C 40 | 215 | 14.7 | 440 | 36 | AA601278 | AA601278 no15b09.s |
| 41 | 214.8 | 14.7 | 624 | 36 | AA601356 | AA601356 no16b05.s |
| C 42 | 214.8 | 14.7 | 623 | 53 | HSM003412 | AI038936 Homo sapi |
| C 43 | 214.6 | 14.7 | 441 | 29 | AA169245 | AA169245 zp19d03.s |
| 44 | 214 | 14.7 | 357 | 45 | AI340832 | AI340832 tb55a10.x |
| C 45 | 214 | 14.7 | 413 | 46 | AI433104 | AI433104 th41d12.x |

ALIGNMENTS

RESULT 1
AI675523/c
LOCUS AI675523 360 bp mRNA
DEFINITION wc01f01.x1 NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2313913 3', mRNA sequence.
ACCESSION AI675523
NID 94876003
VERSION AI675523.1 GI:4876003

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1996 this sequence version replaced gi:1133152.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 357.
Location/Qualifiers
1. 360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone="IMAGE:2313913"
/clone_lib="NCI-CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 89 a 72 c 69 g 130 t
ORIGIN

Query Match 24.6%; Score 358.4; DB 50; Length 360;
Best Local Similarity 99.7%; Pred. No. 3.4e-57;
Matches 359; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 636 GAGAGAACTGAGAGAAACAGAGAAATAACACAGGAATAAGAGCAAGAAGAGA 695
|||||
Db 360 GAGAGAACTGAGAGAAACAGAGAAATAACACAGGAATAAGAGCAAGAAGAGA 301
696 GAAACAGAAACAGACATGGGGAGGAGGAGAAACACACACATAGAAATGCAGTGCCTTC 755
|||||
Db 300 GAAACAGAAACAGACATGGGGAGGAGGAGAAACACACACATAGAAATGCAGTGCCTTC 241
756 CAACAGCATGGGGCTGAGGGCGGTGACCTCCACCCCAATAGAAAATCCTTTAACTTT 815
|||||
Db 240 CAACAGCATGGGGCTGAGGGCGGTGACCTCCACCCCAATAGAAAATCCTTTAACTTT 181
816 TGACTCCCCAAAAACCTGACTAGAAATAGCCTACTGTGACGGGGAGCCTTACCAATAAC 875
|||||
Db 180 TGACTCCCCAAAAACCTGACTAGAAATAGCCTACTGTGACGGGGAGCCTTACCAATAAC 121
876 ATAATAGTCGATTATGCATACGTTTATGCATTCATGATATACCTTTGTGGAATTTT 935
|||||
Db 120 ATAATAGTCGATTATGCATACGTTTATGCATTCATGATATACCTTTGTGGAATTTT 61

QY 936 TTGATATTTCATAGCTACACAGTTGCTGCTGATTTTTTAAATGTGTGCAACTCTCCT 995
|||||
Db 60 TTGATATTTCATAGCTACACAGTTGCTGCTGATTTTTTAAATGTGTGCAACTCTCCT 1

RESULT 2
AI557281 777 bp mRNA EST 23-MAR-1999
LOCUS PT2.1_15.G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION AI557281
ACCESSION 94489644
NID AI557281.1 GI:4489644
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 777)
AUTHORS Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source
1. 777
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6: 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II Kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."

BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 19.6%; Score 286.4; DB 48; Length 777;
Best Local Similarity 79.8%; Pred. No. 6.1e-44;
Matches 387; Conservative 0; Mismatches 14; Indels 84; Gaps 1;

QY 25 AGTAGTGACAGCTCTCTACACCATCGGGCTGGCTGCACAGCTTTGAGGCCGACCAAG 84
||| |
Db 111 ACTGTTTCAGACTCTCTACACCATCGGGCTGGCTGCACAGCTTTGAGGCCGACCAAG 170
85 AGCCAGGAGCCAGATGTGTGAGGGCCAGCTCTCCGTACGGCACCCAGAGTACACAGAC 144
|||||
Db 171 AGCCAGGAGCCAGATGTGTGAGGGCCAGCTCTCCGTACGGCACCCAGAGTACACAGAC 230
145 CCTGCTCGCTAACGACCTCATGCTCATCAAGTGTGACGAATCCGTGTCCGAGTGTGACA 204
|||||
Db 231 CCTGCTCGCTAACGACCTCATGCTCATCAAGTGTGACGAATCCGTGTCCGAGTGTGACA 290
205 CCATCCGGAGCATCAGCATGTGCTTCCGAGTGCCTTACCCGGGGGAACCTTGTGCTCGTTT 264
|||||
Db 291 CCATCCGGAGCATCAGCATGTGCTTCCGAGTGCCTTACCCGGGGGAACCTTGTGCTCGTTT 350
265 CTGGCTGGGGTCTGTGCTGGCAACGGTGAAGCTCAGCGGTGTGTCTGCCCTTTCAAGA 324
|||||
Db 351 CTGGCTGGGGTCTGTGCTGGCAACGG----- 376
325 GGTCTCTGCCAGTGCAGGGGGCTGACCCAGAGCTCTGGCTCCAGCAGAAATGCTTAC 384
|||||
Db 376 -----CAGATGCTAC 386

QY 385 CGTGTGCAGTGCCTGAACGTGTCCGGTGTCTCTGANGAGGCTTCGANTTAAGCTCTATGA 444
|||||
Db 387 CGTGTGCAGTGCCTGAACGTGTCCGGTGTCTCTGAGGAGGCTTCGAGTAAGCTCTATGA 446
QY 445 CCCGCTGTACCACCCCANCATGTCTCTGCGCGCGGAGGAGGCAAGACCAAGAGACTCTCTG 504
|||||
Db 447 CCCGCTGTACCACCCCAGCATGTCTCTGCGCGCGGAGGAGGCAAGACCAAGAGACTCTCTG 506
QY 505 CAACG 509
|||||
Db 507 CAACG 511

RESULT 3
AI686689/c 569 bp mRNA EST 27-MAY-1999
LOCUS tu35g11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253092 3'
DEFINITION similar to TR:Q92046 Q92046 PREPROTRYPsin PRECURSOR ; mRNA
sequence.
ACCESSION AI686689
NID 94897983
VERSION AI686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from G1bco
High quality sequence stop: 444.

FEATURES
source 1..569
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 150 c 170 g 124 t
ORIGIN

Query Match 16.8%; Score 245.4; DB 50; Length 569;
Best Local Similarity 79.6%; Pred. No. 2.3e-36;

Matches 339; Conservative 0; Mismatches 4; Indels 83; Gaps 1;
QY 84 GAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGA 143
|||||
Db 569 GAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGA 510
QY 144 CCCTTGCTGCTTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCCAGTCTGAC 203
|||||
Db 509 CCCTTGCTGCTTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCCAGTCTGAC 450
QY 204 ACCATCCGGAGCATCAGCATTTGCTTCGACAGTGCCCTACCGCGGGGAACCTTGCCTCGT 263
|||||
Db 449 ACCATCCGGAGCATCAGCATTTGCTTCGACAGTGCCCTACCGCGGGGAACCTTGCCTCGT 390
QY 264 TCTGGCTGGGCTCTGCTGGCGAACGGTGAGCTCAGCGGGTGTGTCTGCCCTTCAAGG 323
|||||
Db 389 TCTGGCTGGGCTCTGCTGGCGAAC----- 365
QY 324 AGTCTCTGCCCCAGTCGCGGGGCTGACCCAGAGCTCTGCCGTCCAGCAGATGCCCTA 383
Db 365 -----GGCAGATGCCCTA 353
QY 384 CCGTGTGCAGTGCCTGAACGTGTCCGTTGTGTCTGANGAGGCTTCGANTTAAGCTCTATG 443
|||||
Db 352 CCGTGTGCAGTGCCTGAACGTGTCCGTTGTGTCTGAGGAGGCTTCGAGTAAGCTCTATG 293
QY 444 ACCCGCTGTACCACCCCANCATGTCTCTGCGCGCGGAGGCGCAAGACCAAGAGACTCCT 503
|||||
Db 292 ACCCGCTGTACCACCCCAGCATGTTTTCGCGCGGAGGCGCAAGACCAAGAGACTCCT 233
QY 504 GCAACG 509
|||||
Db 232 GCAACG 227

RESULT 4
AA551449 415 bp mRNA EST 05-SEP-1997
LOCUS nj55e05.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996416
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert length: 640 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 412.

FEATURES
source 1..415
location/Qualifiers


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_lib="NCI_CGAP_Pt9"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."
```

BASE COUNT 78 a 125 c 129 g 83 t

ORIGIN

Query Match 16.8%; Score 244.6; DB 35; Length 415;
Best Local Similarity 78.2%; Pred. NO. 3.3e-36;
Matches 344; Conservative 0; Mismatches 12; Indels 84; Gaps 1;

```
QY 70 TTGAGGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACC 129
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 TCGAGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACC 61
QY 130 CAGAGTACACAGAGCCCTGCTCGCTACGACCTCATGCTCATCAAGTGGAGCAATCCG 189
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 CAGAGTACACAGAGCCCTGCTCGCTACGACCTCATGCTCATCAAGTGGAGCAATCCG 121
QY 190 TGTCCGAGTGTACACCATCCGAGCATCATGCTTCCGAGTGGCCCTACCGCGGGGA 249
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 TGTCCGAGTGTACACCATCCGAGCATCATGCTTCCGAGTGGCCCTACCGCGGGGA 181
QY 250 ACTCTTGCCCTGTTCTGGCTGGGGTCTGCTGGCGAAGGTGAGCTCAGGGGTGTGTCTC 309
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 ACTCTTGCCCTGTTCTGGCTGGGGTCTGCTGGCGAAGGTGAGCTCAGGGGTGTGTCTC 222
QY 310 TGGCCTCTTCAAGAGGTCTCTTGGCCAGTGGCGGGGCTGACCCAGAGCTTGTCCC 369
Db 222 -----
QY 370 AGGCAGATGCTTACCGTGTGAGTGCAGTGCAGTGAACGTGTGGTGTCTGANGAGTCTG 429
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 ----CAGATGCTTACCGTGTGAGTGCAGTGCAGTGAACGTGTGGTGTCTGANGAGTCTG 277
QY 430 CANTAACTCTATGACCCGCTGTACCAACCCANCAATGTTCTGCGCCGCGGAGGCAAGA 489
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 CAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGCAAGA 337
QY 490 CCAGAAGGACTCTTCAACG 509
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 CCAGAAGGACTCTTCAACG 357
```

RESULT 5
LOCUS AI557025 722 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557025
NID g4489388
VERSION AI557025.1 GI:4489388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Huang,G.M., Ng,W., Parkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
YU,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

Contact: Guyang Matthew Huang

Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangym@yahoo.com.
Location/Qualifiers

FEATURES
source 1. 722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15: 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."

BASE COUNT 146 a 156 c 195 g 176 t 49 others

ORIGIN

Query Match 16.6%; Score 242.8; DB 48; Length 722;
Best Local Similarity 90.0%; Pred. NO. 6.7e-36;
Matches 298; Conservative 0; Mismatches 25; Indels 8; Gaps 4;

```
QY 36 AGCTCTACACCATCGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 95
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AGCTTCTACACCATCGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 174
QY 96 CAGATGTGAGGCCAGCCTCTCCGTACGGCCACCCAGAGTACACAGACCCTTGTGCT 155
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 CAGATGTGAGGCCAGCCTCTCCGTACGGCCACCCAGAGTACACAGACCCTTGTGCT 234
QY 156 AACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGC 215
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 AACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGC 294
QY 216 ATCAGCATTGTCTTGCAGTGCCTTACCCGCGGGGAACCTTGTCC-TGGTTCTGGCTGGGG 274
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 ATCAGCATTGTCTTGCAGTGCCTTACCCGCGGGGAACCTTGTCC-TGGTTCTGGCTGGGG 354
QY 275 TCTGC-TGGCGAAGGTGAGCTCAGCGGTGTGTGTCTGCCCTCTTC---AAGAGGTCC 329
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 TCTGCTTGGCGAAGGTGAGCTCAGCGGTGTGTGTCTGCCCTTNTTAAAGAGAGTCC 414
QY 330 TCTGCCCA--GTGCGGGGGGCTGACCCAGAG 358
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 TTTTNCNAAGTNTGGGGGGCTNACCCAGAG 445
```

RESULT 6
LOCUS AI669421 481 bp mRNA EST 14-MAY-1999
DEFINITION ty32d03.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2280773 3',
similar to contains Alu repetitive element; mRNA sequence.
ACCESSION AI669421
NID g4834195
VERSION AI669421.1 GI:4834195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1130773.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 425.

FEATURES
source location/Qualifiers

1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2280773"
/clone_1lb="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 114 a 119 c 106 g 142 t
ORIGIN

Query Match 15.5%; Score 226.2; DB 50; Length 481;
Best Local Similarity 81.1%; Pred. No. 8.1e-33;
Matches 275; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 1122 ACACGAGAGAAACAGGAAAAATCAAGACTCTACAAAGAGGCTGGCAGGCTGCATG 1181
||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 339 ACTCTATAGAATCTATAAATTTTATTATTATAAATAGCAGCCGGCATGTGATCAGC 280

QY 1182 CCTGTATCCAGCACTTTGGAGG-CGAGGCGAGCAGATCATTGAGGTAAGAGTTCA 1240
|||| | ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 279 CCTGTGTTCCAGCACTTTGGAGGCGCAGGCGAGCAGATCATTGAGGTCAGAGTTTG 220

QY 1241 AGACGAGCCTGGCCAAATGTGAATCCTGCTGTACTATAAAATACAAAAGTTAGCTGG 1300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 219 AGACGAGCCTGGCCAAATGTGAATCCTGCTGTACTATAAAATACAAAATTTAGCTGG 160

QY 1301 ATATGCTGCGAGCGCCTGTAATCCAGCTACTTGGAGGCTGAGGCGAGGAATTGCTT 1360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 159 GCGTGTGGCAGGCGCCTGTAACCCAGCTACTTGGAGGCTGAGGCGAGGAATTGCTT 100

QY 1361 GAATATGGAGGCGAGGTTGAAGTGAGTTGAGATCACACCACATTAATCCAGCTGGGC 1420
|| | ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db 99 GACTCTGGAGGCGAGGTTGACGTAGCAGCAAGATTGGCCACTGCACCTCAGCCTGGGT 40

QY 1421 AACAGAGTAAGACTCTGCTCAAAAAAATAAAAAA 1459
|||| | ||||||| ||| ||||||| |||||||
Db 39 GACAGGGTGAGACTCCATCTCTAAAAAACAATAAAGA 1

RESULT 7
AI609972 415 bp mRNA EST 21-APR-1999
LOCUS tt78c10.x1 NCI-CGAP_HSC3 Homo sapiens cDNA clone IMAGE:2246898 3'
DEFINITION similar to contains Alu repetitive element; contains element MER9
repetitive element ;, mRNA sequence.
ACCESSION AI609972
NID 94619139
VERSION AI609972.1 GI:4619139
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948944.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

FEATURES
source location/Qualifiers

1. .415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2246898"
/clone_1lb="NCI-CGAP_HSC3"
/tissue_type="CD34+, T negative, patient with chronic myelogenous leukemia"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: PAMPI; mRNA made from lymphoid tissue, CDNA made by oligo-dT priming. directionally cloned. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified. CDNA Library Preparation: David B. Krizman, Ph.D. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 132 a 81 c 125 g 77 t
ORIGIN

Query Match 15.4%; Score 225.4; DB 48; Length 415;
Best Local Similarity 87.6%; Pred. No. 1.2e-32;
Matches 269; Conservative 0; Mismatches 36; Indels 2; Gaps 2;

QY 1154 ACAAGAGGCTGGGCGAGGCTGCATGCTGTATATCCAGCACTTTGGAGG-CGAGGC 1212
| | ||||||| ||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 91 AAAGAGGCTGGGCATGTGTCTCTGCTGTATCCAGCACTTTGGAGGCGAGGT 150

QY 1213 AGCAGATCACTTGAGGTAAGAGTTCAAGACCAGCCTGGCCAAATGTTGAATCCTGT 1272
||| ||| | ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 151 GGGCGATTAATCTGAGTCAAGAGTTCAAGACCAGCCTGGCCAAATGTTGAATCCTGT 210

QY 1273 CTGTACTAAATAACAAAGTTAGCTGATATGTTGGCAGGCGCCTGTATCCAGCTAC 1332
|| ||||||| ||||||| |||| | | ||||||| ||||||| ||||||| ||||||| |||
Db 211 CTCTACTAAATAACAAATTTAGCCGGGTGTGTGGCATGCGCCTGTATCCAGCTAC 270

QY 1333 TTGGAGGCTGAGGCGAGGAGATTGCTTGAATATGGGAGGCGAGAGTTGAAGTGAAGTTGA 1392
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db 271 TTGGAGGCTGAGGCGAGGAGATTGCTTGAACCTGGAGGC-GAGGTTGACAGTGAAGCCGA 329

QY 1393 GATCACACCACTATATCTCCAGCTGGGGCAACAGAGTTAAGACTGTCTCAAAAAA 1452
||||| ||||||| ||||||| ||| | | ||||||| ||||||| ||||||| |||||||
Db 330 GATCACACCACTGCACTCCAGCCTGGATGAATAACAGCGAGACTGTCTCAAAAAA 389

QY 1453 AAAAAA 1459
|||||
Db 390 AAAAAA 396

RESULT 8
AA535216 407 bp mRNA EST 21-AUG-1997
LOCUS AA535216
DEFINITION nj75f05.s1 NCI-CGAP_P10 Homo sapiens cDNA clone IMAGE:998337
similar to contains Alu repetitive element; contains element MER22
MER22 repetitive element ;, mRNA sequence.

ACCESSION AA535216
NID 92279469
VERSION AA535216.1 GI:2279469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 407)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On May 5, 1995 this sequence version replaced gi:798604.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquil,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
source Location/Qualifiers

1..407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone="IMAGE:998337"
/clone_lib="NCI_CGAP_Pr10"
/sex="male"
/tissue_type="Invasive prostate tumor"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
invasive prostate tumor, CDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

BASE COUNT 112 a 93 c 123 g 79 t
ORIGIN

Query Match 15.3%; Score 223; DB 35; Length 407;
Best Local Similarity 86.3%; Pred. No. 3.2e-32;
Matches 258; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

OY 1161 GGCTGGGAGGGTGGCTCATGCGCTGTAATCCAGCACTTTGGAGGC-GAGGCAGGCAGA 1219
|||||
DB 107 GGCTGGGAGGGTGGCTCATGCGCTGTAATCCAGCACTTTGAGAGGCTAAGGCAGGCAGA 166
OY 1220 TCACCTGAGGTAAGGAGTTCAAGACCAAGCCTGGCCAAATGGTGAATCCTGCTGTACT 1279
|||
DB 167 TCGCCTGAGTCCAGGGGTTCAAGACCAAGCCTGGCCAAACATGTGAAACCCCGTCTTACT 226
OY 1280 AAAAATACAAAAGTTAGCTGATATGTTGGCAGGCGCCTGTAATCCAGCACTTGGGAG 1339
|||||
DB 227 AAAAATACAAAAGTTAGCTGAGCAGGTTGGCAGCTGCTGTAATCCAGCACTTGGGAG 286
OY 1340 GCTGAGCAGAGAAATGCTGTAATGGAGCAGAGAGTTGAAGTGAAGTGAATCACA 1399
|||||
DB 287 GCTGAGCAGAGAAATGCTGTAATCCAGAGAGCGGAAGTTGCAAGTGAAGCAGATCGTG 346
OY 1400 CCACATTAATCCAGCTGGGAGCAAGAGTAAGACTCTCTCAAAAAAAAAAAAAAAAAA 1458
|||||
DB 347 CCACATTAATCCAGCTGGGAGCAAGAGTAAGACTCTCTCAAAAAAAAAAAAAAAAAA 405

RESULT. 9 - .1

W96522/c
LOCUS W96522 479 bp mRNA EST 29-NOV-1996
DEFINITION ze43f08.r1 Soares retina N2b4HR Homo sapiens CDNA clone
IMAGE:361767 5' similar to contains Alu repetitive element; mRNA
sequence.

ACCESSION W96522
NID 91426448
VERSION W96522.1 GI:1426448
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 479)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL On Nov 29, 1993 this sequence version replaced gi:636214.
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3253 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 457.

FEATURES
source Location/Qualifiers

1..479
/organism="Homo sapiens"
/db_xref="GDB:1278470"
/db_xref="taxon:9606"
/map="1"
/clone="IMAGE:361767"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand CDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGGCGGCCGCTTTTCTTTTCTTTT 3'],
double-stranded CDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 89 a 134 c 100 g 152 t 4 others
ORIGIN

Query Match 15.2%; Score 222.4; DB 26; Length 479;
Best Local Similarity 83.3%; Pred. No. 4.1e-32;
Matches 264; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

OY 1144 TCAGACTCTACAAGAGGCTGGGAGGGTGCCTCATGCGCTGTAATCCAGCACTTTGGG 1203
|||||
DB 338 TCAAAAAGATAAAGGAGCGCGGAGGACAGTGCCTCATGCGCTGTAATCCAGCACTTTGGG 279
OY 1204 AGGC-GAGGAGGACATCACTTGAGGTAAGAGGTTCAAGACCAAGCCTGGCCAAATGGT 1262
|||||

Db 278 AGGCTGAGTGGGCGAGATCACCCTGAGTTCAGAGAGCTGGCCACGCTGGT 219
QY 1263 GAAATCCTGCTGTACTATAAATAAGAGTAGCTGATATGCTGGCAGCGCCTGTAA 1322
Db 218 GAAACCCCGCTCTACTAAANATACAAAATTAAGTGGTATGCTAGTGTCTCTCTAA 159
QY 1323 TCCAGCTACTTGGAGGCTGAGGCGAGAAATTGCTGAATATGGGACAGAGGTTGA 1382
Db 158 TCCAGCTACTTGGAGGCTGAGGCGAGAAATTCATGAACCGAGGAGGAGGTTGC 99
QY 1383 AGTGAGTGAATTCACACCACTATATCTCCAGCTGGGCAACAGAGTAAGACTCTGTCTCA 1442
Db 98 AGTGAGCGGAGATCACACCACTGCCTGACCGCTGGACAACAGAGTAAGACTCTGTCTAA 39
QY 1443 AAAAAAAAAAAAAA 1459
Db 38 AAAAAATAACAATAA 22

RESULT 10
AA644090/c 437 bp mRNA EST 06-MAR-1998
LOCUS ab62b04.s1 Stratagene lung carcinoma 937218 Homo sapiens cDNA clone
DEFINITION IMAGE:845359 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION AA644090
NID g2569308
VERSION AA644090.1 GI:2569308
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 437)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
TITLE On Sep 19, 1997 this sequence version replaced gi:1520385.
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 834 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 416.

FEATURES

source

1. 437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:845359"
/clone_1lb="Stratagene lung carcinoma 937218"
/tissue_type="lung carcinoma"
/cell_line="NCI-H69"
/dev_stage="cell line NCI-H69"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Small cell carcinoma cell line NCI-H69. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3'."
BASE COUNT 98 a 112 c 78 g 149 t
ORIGIN

Query Match 15.2%; Score 222.2; DB 36; Length 437;
Best Local Similarity 85.5%; Pred. No. 4.5e-32;
Matches 259; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
QY 1158 AGAGCTGGCGAGGCTGCTCATGCTGTAAATCCAGCACTTTGGAGG-CGAGCGAGGC 1216
Db 310 ACAGGCTGGGTACAGGGGCTCATGCTGTAAATCCAGCACTTTGGAGGCGGCGGGC 251
QY 1217 AGATCACTTGAGGTAAGAGTTTCAAGACCAAGCCCTGGCCAAATGCTGAATCTGTCTGT 1276
Db 250 AGATCACTTGAGATCAGAGAGTTGAGACCAAGCCCTGGCCAAATGCTGAATCTGTCTGT 191
QY 1277 ACTAAAAATACAAAAGTTAGCTGATATGCTGGCAGGCGCCTGTAAATCCAGCACTTGG 1336
Db 190 ACTAAAAATACAAAATTAGCCAGGTGTGTGGCAGGTGCTGTAAATCCAGCACTTCTGG 131
QY 1337 GAGCTGAGCGCAGAGAAATTGCTGAATATGGGAGCGAGGTTGAAGTGAGTTGAGATC 1396
Db 130 GAGCTGAGCGCAGAGAAATCGTTGAACCCGGAGGTAAGGTTGCAGTGAGCCAGATC 71
QY 1397 ACACCACTAATCTCCAGCTGGGCGAACAGAGTAAGACTCTGTCAAAAAAAAAAAAAA 1456
Db 70 GCACCACTGCATCCAGCCTGGGTGATAGAGTAAGACTCAGTCTCAAAAAAAAAAAAAA 11
QY 1457 AAA 1459
Db 10 AGA 8

RESULT 11
AA626040/c 500 bp mRNA EST 15-OCT-1997
LOCUS af94a08.s1 Soares total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:1055414 3' similar to contains Alu repetitive
element;contains element MER22 repetitive element;; mRNA sequence.
ACCESSION AA626040
NID g2538427
VERSION AA626040.1 GI:2538427
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 500)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Giesel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
TITLE On Apr 14, 1993 this sequence version replaced gi:503350.
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES

source

1. 500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1055414"
/clone_1lb="Soares total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dt) primer [5'

| BASE COUNT | ORIGIN |
|------------|--|
| 115 a | TGTTACCAATCTGAAGTGGAGCGCGCCGCTAATTTTTTTTTTTTTTTT 3'] . |
| 123 c | Double-stranded cDNA was ligated to Eco RI adaptors |
| 107 g | (Pharmacia), digested with Not I and cloned into the Not I |
| 154 t | and Eco RI sites of the modified pT73 vector. Library |
| 1 others | went through one round of normalization, and was |
| | constructed by Bento Soares and M. Fatima Bonaldo. " |

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match | 15.18; | Score 220.2; | DB 36; | Length 500; |
| Best Local Similarity | 81.28; | Pred. No. 1e-31; | | |
| Matches 255; Conservative | 0; | Mismatches 59; | Indels 0; | Gaps 0; |

[illegible]

| | | | | |
|------------|--|--------------------------------------|-------------------------|----------------------------|
| RESULT | 12 | | | |
| LOCUS | AI634187/c | | | |
| DEFINITION | AI634187 | 454 bp | mRNA | EST |
| | ts55a08.x1 | NCI_CGAP_R1d8 | Homo sapiens | CDNA clone IMAGE:2232470 3 |
| | similar to | contains | Alu repetitive element; | contains element TARI |
| | repetitive | element ; | mRNA | sequence. |
| ACCESSION | AI634187 | | | |
| NID | q4685517 | | | |
| VERSION | AI634187.1 | GI:4685517 | | |
| KEYWORDS | EST. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | |
| | Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| REFERENCE | 1 (bases 1 to 454) | | | |
| AUTHORS | NCI-CGAP | http://www.ncbi.nlm.nih.gov/ncicgap. | | |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), | | | |
| JOURNAL | Tumor Gene Index | | | |
| COMMENT | Unpublished (1997) | | | |
| | On May 18, 1998 this sequence version replaced gi:3138239. | | | |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-dio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

| FEATURES | |
|----------|-----------------------------------|
| source | High quality sequence stop: 411.. |
| | Location/Qualifiers |
| | 1. .454 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5"
/clone="IMAGE:2232470"
/clone_lib="NCI-CGAP_K1d8"
/tissue_type="renal cell tumor"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.2 kb. Life Technologies catalog #:
11524-014"

```

| BASE COUNT | 118 a | 103 c | 93 g | 140 t | ORIGIN |
|-----------------------|---|------------------|-----------|-------------|--------|
| Query Match | 15.1%; | Score 220.2; | DB 49; | Length 454; | |
| Best Local Similarity | 81.8%; | Pred. No. 1e-31; | | | |
| Matches 266; | Conservative 0; | Mismatches 58; | Indels 1; | Gaps 1; | |
| OY 1130 | AGAAACAGGAAAAATCAAGACTCTACAAGAGGCTGGGACGGCTGCATGCGCTGTAAT | 1189 | | | |
| Db 325 | AGTAAATTATATATAATAAAATACCAAAAAGAGGCTGGGCTGGCTCACGCGCTGTAAT | 266 | | | |
| OY 1190 | CCGACACATTGGGAGG-CGAGGCGAGGCGAGATCATTGAGGTAAGAGATTCAAGACCAGC | 1248 | | | |
| Db 265 | CCCAGCATTGGGAGGCGCGAGGCGAGGCTGGATCACCCTAAGCTCAGAGATTGAGACCAGC | 206 | | | |
| OY 1249 | CTGGCCAAATGGTGAATCCTGCTCTGACTAAAAATACAAAAGTTAGCTGATATGGTG | 1308 | | | |
| Db 205 | CTGGCCAAATATGTTGAACCCCATCTCTACTAAAAATACAAAATTAGCCGGGTGGTG | 146 | | | |
| OY 1309 | GCAGGCGCCTGTATCCACGCTACTTGGGAGGCTGAGGCGAGGAGAATTGCTGGAATATGG | 1368 | | | |
| Db 145 | GCGTGCCTGTAGTCCACGCTACTCGGAGGCTGAAGCAGAGAATCGCTGAACCCGG | 86 | | | |
| OY 1369 | GAGGCGAGAGTTGAAGTGAATTGAGATCACACCCTATATTCTCCAGCTGGGGCAACAGAGT | 1428 | | | |
| Db 85 | GAGGCGGAGGTTACAGTGAAGTGAAGATTGTGCCACTGCACCTGGGTGACAGAGT | 26 | | | |
| OY 1429 | AAGACTCTGTCTCAAAAAAAAAAAAA 1453 | | | | |
| Db 25 | GAGATTCCGCTCAAAAAAAAAAAAA 1 | | | | |

| | |
|-----------------|---|
| RESULT | 13 |
| LOCUS | AA746911 |
| DEFINITION | AA746911 374 bp mRNA EST 22-JAN-1998 nx69c10.s1 NCI_CGAP_Alvl Homo sapiens cDNA clone IMAGE:1267506 similar to contains Alu repetitive element;; mRNA sequence. |
| ACCESSION | AA746911 |
| NID | g2786869 |
| VERSION | AA746911.1 GI:2786869 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 374) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) |
| JOURNAL COMMENT | On Dec 12, 1995 this sequence version replaced gi:1119179. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA library Preparation: David B. Krizman, Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.

This Page Blank (uspto)

OM of: US-09-030-606-175 to: A_Geneseq_36:* out_format : pfs
Date: Sep 25, 1999 11:35 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=framed_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29804/app_query.fasta.1
-DB=A_Geneseq_36-QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-175
Query length: 1167
Database: A_Geneseq_36:*
Database sequences: 188963
Database length: 23686106
Search time (sec): 185.540000

| Sequence | Strd Orig | ZScore | Escore | len | Documentation |
|---------------------|-----------|---------|---------|-----|---------------------------------|
| A_Geneseq_36:W59129 | 1113.00 | 1278.98 | 3.9e-64 | 232 | Homo sapiens Tub Interactor (h |
| A_Geneseq_36:W60592 | 1091.00 | 1253.25 | 9.9e-63 | 248 | Human prostate-specific kallikr |
| A_Geneseq_36:W69388 | 1087.00 | 1250.17 | 1.8e-62 | 205 | Prostate tumour specific gene c |
| A_Geneseq_36:W71872 | 1087.00 | 1250.17 | 1.8e-62 | 205 | Protein encoded by prostate tum |
| A_Geneseq_36:W69387 | 807.00 | 931.45 | 1.3e-44 | 159 | Prostate tumour specific gene c |
| A_Geneseq_36:W71871 | 807.00 | 931.45 | 1.3e-44 | 159 | Protein encoded by prostate tum |
| A_Geneseq_36:W69389 | 566.00 | 655.15 | 3.1e-29 | 164 | Prostate tumour specific gene c |
| A_Geneseq_36:W71873 | 566.00 | 655.15 | 3.1e-29 | 164 | Protein encoded by prostate tum |
| A_Geneseq_36:W94493 | 553.00 | 636.38 | 2.1e-28 | 268 | Human kallikrein. Human kallikr |
| A_Geneseq_36:R67888 | 530.50 | 611.07 | 5.7e-27 | 253 | Human stratum corneum chymotrop |
| A_Geneseq_36:W05383 | 530.50 | 611.07 | 5.7e-27 | 253 | Human amyloid precursor protein |
| A_Geneseq_36:W10694 | 498.50 | 574.20 | 6.3e-25 | 260 | Mouse recombinant neurotrophin |
| A_Geneseq_36:W12393 | 498.50 | 574.20 | 6.3e-25 | 260 | Human amyloid precursor protein |
| A_Geneseq_36:R44532 | 470.50 | 542.62 | 3.9e-23 | 244 | Zyme APP-cleaving protease. Amy |
| A_Geneseq_36:W22985 | 470.50 | 542.62 | 3.9e-23 | 244 | Human serine protease 59 (SP59) |
| A_Geneseq_36:W51006 | 470.50 | 542.62 | 3.9e-23 | 244 | Human serine protease 59 (SP59) |
| A_Geneseq_36:W87703 | 466.50 | 537.54 | 6.9e-23 | 260 | A human serine protease designa |
| A_Geneseq_36:R53638 | 457.50 | 528.20 | 2.6e-22 | 230 | Bovine trypsinogen. Expression |
| A_Geneseq_36:W81767 | 456.50 | 527.30 | 3.0e-22 | 223 | Bovine trypsin. Expression vect |
| A_Geneseq_36:R53637 | 456.50 | 527.30 | 3.0e-22 | 224 | Bovine trypsin. Expression vect |
| A_Geneseq_36:W45398 | 450.00 | 519.37 | 7.8e-22 | 237 | Prostate-specific antigen prote |
| A_Geneseq_36:W45395 | 449.00 | 518.23 | 9.1e-22 | 237 | Human prostate-specific antigen |
| A_Geneseq_36:W56086 | 449.00 | 518.23 | 9.1e-22 | 237 | Human prostate-specific antigen |
| A_Geneseq_36:W83213 | 449.00 | 518.23 | 9.1e-22 | 237 | Prostate-specific antigen prote |
| A_Geneseq_36:W83202 | 449.00 | 518.23 | 9.1e-22 | 237 | Prostate-specific antigen prote |
| A_Geneseq_36:W96186 | 449.00 | 518.23 | 9.1e-22 | 237 | Human prostate-specific antigen |
| A_Geneseq_36:W96187 | 449.00 | 518.23 | 9.1e-22 | 237 | Human prostate-specific antigen |
| A_Geneseq_36:W45396 | 449.00 | 518.00 | 9.1e-22 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W83204 | 449.00 | 518.00 | 9.1e-22 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W96188 | 449.00 | 518.00 | 9.1e-22 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W06971 | 449.00 | 517.46 | 9.1e-22 | 261 | Human prostate-specific antigen |
| A_Geneseq_36:W45397 | 449.00 | 517.46 | 9.1e-22 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W49085 | 449.00 | 517.46 | 9.1e-22 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W83203 | 449.00 | 517.46 | 9.1e-22 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W96189 | 449.00 | 517.46 | 9.1e-22 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:R84669 | 448.00 | 516.85 | 1.1e-21 | 244 | Pro-hk2 kallikrein. New isolate |
| A_Geneseq_36:R84668 | 448.00 | 516.32 | 1.1e-21 | 261 | Prepro-hk2 kallikrein. New isol |
| A_Geneseq_36:R84667 | 447.00 | 515.94 | 1.2e-21 | 237 | Mature kallikrein hk2. New isol |
| A_Geneseq_36:W10600 | 447.00 | 515.17 | 1.2e-21 | 261 | Human prepro-tryp26-glandular k |
| A_Geneseq_36:W49087 | 446.00 | 514.79 | 1.4e-21 | 237 | Mutant human kallikrein 2 (hk2) |
| A_Geneseq_36:W83212 | 446.00 | 514.79 | 1.4e-21 | 237 | hk2 variant A217V. Detection of |

| | | | |
|--|---|--------------------------|--|
| seq_name: A_Geneseq_36:W59129 | | seq_documentation_block: | |
| ID | W59129 standard; Protein; 232 AA. | | |
| AC | W59129; | | |
| DT | 11-SEP-1998 (first entry) | | |
| DE | Homo sapiens Tub Interactor (hTI-1) protein. | | |
| KW | serine protease; tub interactor; treatment; obesity; cachexia; | | |
| KW | anorexia nervosa; diabetes; cell cycle progression; apoptosis; | | |
| KW | neurodegenerative disease; Alzheimer's disease; drug screening; | | |
| KW | Parkinson's disease; Huntington's chorea; detection; diagnosis; | | |
| KW | amyotrophic lateral sclerosis; spinocerebellar degeneration. | | |
| OS | Homo sapiens. | | |
| FH | Key | Location/Qualifiers | |
| FT | Region | 42 | |
| FT | /note="undefined amino acid" | | |
| PN | W09812302-A1. | | |
| PD | 26-MAR-1998. | | |
| PF | 05-SEP-1997; U15627. | | |
| PR | 21-JUL-1997; US-897340. | | |
| PR | 17-SEP-1996; US-715032. | | |
| PA | (MILL-) MILLENNIUM PHARM INC. | | |
| PI | Errada PR, Gimeno CJ; | | |
| DR | WPI; 98-217246/19. | | |
| DR | N-PSDB; V11855. | | |
| PT | Tub interactor genes - used to develop products for the treatment | | |
| PT | of obesity, cachexia, anorexia nervosa or related disorders e.g. | | |
| PT | diabetes | | |
| PS | Claim 28; Fig 1; 120pp; English. | | |
| CC | The sequence is that encoding the Tub Interactor protein (hTI-1) | | |
| CC | which is a putative serine protease. TI genes function | | |
| CC | in biochemical pathways involved in weight control and | | |
| CC | related disorders. The products can be used for treating | | |
| CC | weight disorders, e.g. obesity, cachexia or anorexia nervosa, | | |
| CC | or a related disorder such as diabetes. The products can | | |
| CC | also be used to modulate cell cycle progression and apoptosis. | | |
| CC | They can be used for treating neurodegenerative diseases | | |
| CC | which are characterised by apoptosis, including Alzheimer's | | |
| CC | disease, Parkinson's disease, Huntington's chorea, amyotrophic | | |
| CC | lateral sclerosis or spinocerebellar degenerations. The | | |
| CC | products can also be used for detection, diagnosis and | | |
| CC | drug screening. | | |
| SO | Sequence 232 AA; | | |
| alignment_scores: | | | |
| Quality: 1116.00 | | Length: 226 | |
| Ratio: 5.096 | | Gaps: 1 | |
| Percent Similarity: 96.903 | | Percent Identity: 95.133 | |
| alignment_block: | | | |
| US-09-030-606-175 x W59129 .. | | | |
| Align seg 1/1 to: W59129 from: 1 to: 232 | | | |
| 1 | CGCGAGCCCTGGCAGGGCGGACATGTCATGCAAAAGCAATGTTCTGCTC | 50 | |
| : | ::: | | |
| 7 | SergInProTprglnAlaAlaLeuValMetGluAsnGluLeuPheCysSe | 23 | |
| : | | | |
| 51 | GGGCTCTGCTGTCATCCGACAGTGGTGTGTACAGCCGACACTGTTTC | 99 | |
| : | | | |
| 23 | rglyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysPheG | 40 | |
| : | | | |
| 100 |CAGAATCTCTACACATCGGGCTGGGCTGCACAGTCTT | 138 | |
| : | | | |
| 40 | Inlys**ValGlnSerSerTyrThrIleGlyLeuGlyLeuHisSerLeu | 56 | |
| : | | | |
| 139 | GAGCGCAGCAGAGCCAGGAGGAGGAGATGTGTGAGGCCAGCCTCTCCGT | 188 | |
| : | | | |

57 GluAlaAspGlnGluProGlySerGlnMetValGluAlaSerLeuSerVa 73
189 ACGGACACCAGAGTACACAGACCTCTGCTACGCTACGACCTCATGCTCA 238
|||||
73 LArgHisProGlnuTyraSnaArgProLeuLeuAlaAsnAspLeuMetLeuI 90
239 TCAAGTTGAGCAATCCGCTGTCCGAGTCTGACACCATCCGAGACATCAGC 288
|||||
90 lElYsLeuAspGlnuSerValSerGlnuSerAspThrIleArgSerIleSer 106
289 ATGCTTCGACAGTGCCTTACCCGGGGGAACCTTGCCCTCGTNTCTGGCTG 338
|||||
107 lIeAlaSerGlnCysProThrAlaGlnAsnSerCysLeuValSerGlyTr 123
339 GGGTCTGTGGCGGAACGGCAGAAATGCTTACCGTGTGCTGACCTGCGTAACG 388
|||||
123 pGlyLeuLeuAlaAsnGlnYArgMetProThrValLeuGlnCysValAsnV 140
389 TGTGGTGGTGTCTGAGANGTCTGCAAGTAAGCTCTATGACCCGCTGTAC 438
|||||
140 aISerValValSerGlnuGluValCysSerLysLeuTyraSPProLeuTy 156
439 CACCCAGCATGTCTTCCGCGCGCGGAGGGCAAGACCAGAAGACTCCTG 488
|||||
157 HisProSerMetPheCysAlaGlnGlyGlnGlnAspGlnLysAspSercY 173
489 CAACGGTGACTCTGGGGGGCCCGCTGATCTGCAACGGGTACTTGCAAGGCC 538
|||||
173 sAsnGlnYAspSerGlnGlyProLeuIleCysAsnGlnYTrLeuGlnGlnYL 190
539 TTGTGTCTTTCGGAAGAACCCCGCTGTGGCCCACTGGCGCTGAGGTGTC 588
|||||
190 euValSerPheGlnLYLysAlaProCysGlnGlnValGlyValProGlyVal 206
589 TACACCAACCTCTGCAAAATTCACTGAGTGAGATAGAGAAACCGT.CCAGN 637
|||||
207 TyrThrAsnLeuCysLysPheThrGlnuTrpIleGlnuLysThrValProG 223
638 CCAGTTAACTCTGGGAGACTGGAAACCA 665
|||||
223 yGlnLeuThrLeuGlnYThrGlnYAsnPro 232

seq_name: A_Geneseq_36:W60592

seq_documentation_block:

ID W60592 standard; Protein; 248 AA.
AC W60592;
DT 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) protein.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK.
OS Homo sapiens.
FH key location/Qualifiers
FT Misc_difference 113 /label= unknown
FT /note= "encoded by NTC"
FT Misc_difference 128 /label= unknown
FT /note= "unknown"
FT Misc_difference 132 /note= "encoded by AGN"
FT /label= unknown
FT /note= "unknown"
FT /label= "encoded by GNT"
PN WO9820117-A1.
PD 14-MAY-1998.
PR 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goll SK;
DR WPI: 98-286933/25.
DR N-PSDB: V37495.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia

PS Claim 1; Fig 1A-C; 68bp; English.
CC This represents a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SO Sequence 248 AA;

alignment_scores:
Quality: 1091.00 Length: 214
Ratio: 5.195 Gaps: 0
Percent Similarity: 98.131 Percent Identity: 94.860

alignment_block:
US-09-030-606-175 x W60592 ..

Align seg 1/1 to: W60592 from: 1 to: 248

1 GCGCAGCCCTGGCAGGGCGGCACTGTCATGGAAGAAATGTTCTGCTC 50
:::
35 SerGlnProTrpGlnAlaAlaLeuValMetGlnuSnglnuLeuPheCysSe 51
51 GGGCGTCTGTGCATCCGAGTGGGTGCTGTGACCGCACACTGTTCC 100
|||||
51 rGlyValLeuValHisProGlnuTrpValLeuSerAlaAlaHisCysPheG 68
101 AGAATCCTTACACCATCGGCTGGGCTGACACAGTCTTGAGCCGACCAA 150
|||||
68 lnsAsnSerTyTrThrIleGlnYLeuGlnYLeuHisSerLeuGluAlaAspGln 84
151 GAGCCAGGAGCCAGATGGTGGAGGCCAGCCCTCCCTACGGCACCCAGA 200
|||||
85 GluProGlySerGlnMetValGluAlaSerLeuSerValArgHisProG 101
201 GTACAACAGACTCTTGCTCCGCTTAACGACCTCATGCTCATCAAGTTGACG 250
|||||
101 uTyraSnaArgProLeuLeuAlaAsnAspLeuMet***IleLysLeuAspG 118
251 AATCCGTGCTCCGAGTCTGACACCATCCGAGCATCAGCATTTGCTTCGAG 300
|||||
118 luserValSerGlnuSerAspAsnIleArg***IleSerIle***SerGln 134
301 TGCCCTACCCGGGGAACTTGTGCTCGTNTCTGCTGGGCTGTGCTGGC 350
|||||
135 CysProThrAlaGlnYAsnPheCysLeuValSerGlyTyPglyLeuLeuAl 151
351 GAACGGCAGAAATGCCCTACCGTCTGCACTGCGTGAACGTGTGCGTGT 400
|||||
151 aAsnGlnYArgMetProThrValLeuGlnCysValAsnValSerValValS 168
401 CTGAGGANGTCTGCAAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATG 450
|||||
168 erGlnuGlnValCysSerLysLeuTyraSPProLeuTyTrHisProSerMet 184
451 TTCTGCGCGCGGGGAGGGCAAGACCAGAAGAGACTCCTGCAACGGGTGACTC 500
|||||
185 PheCysAlaGlnGlyGlnGlnAspGlnLysAspSercYAsnGlnYAspSe 201
501 TGGGGGCCCCGTGATCTGCAACGGGTACTTGACAGGCCCTTGTGCTTTTCG 550
|||||
201 rGlnGlyProLeuIleCysAsnGlnYTrLeuGlnGlnYLeuValSerPheG 218
551 GAAAAGCCCCGTGTGCCAACTGGCGTGCAGAGTGTCTACACCAACCTC 600
|||||
218 lYlYsAlaProCysGlnGlnValGlyValProGlnYValTyTrThrAsnLeu 234
601 TGCAATTCACCTGAGTGGATAGAGAAACCGCTCCAGNCCAGT 642

|||||
235 CysLysPheThrGluTrpIleGlyLysThrValGlnAlaSer 248

seq_name: A_Geneseq_36:W69388

seq_documentation_block:

ID W69388 standard; Protein; 205 AA.

AC W69388;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DE13 protein.

KW Prostate tumour specific gene; human; prostate cancer; detection; therapy.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 127 /note- "unspecified amino acid"

FT Misc_difference 204 /note- "unspecified amino acid"

FT W09837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI: 98-480805/41.

DR N-PSDB; V58647.

PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers

PS Example 1; Page 115-116; 141pp; English.

CC This sequence is encoded by a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as

CC this protein sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate cancers.

CC Sequence 205 AA;

SO

alignment_scores: Quality: 1087.00 Length: 205
Ratio: 5.355 Gaps: 0
Percent Similarity: 99.024 Percent Identity: 99.024

alignment_block:

US-09-030-606-175 x W69388

Align seg 1/1 to: W69388 from: 1 to: 205

28 ATGGAACGAATTGTTCTGCTGGCGCTCTGTCATCCGACGTGGGT 77
|||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
78 GCTGTACGCCGACACTGTTCCAGAACTCTACACCATCGGGCTGGCC 127
|||||
17 IleuSerAlaIleHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
128 TGACACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGCC 177
|||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
178 AGCCTTCCTGACGGACACCCAGAGTACACAGACTCTTGCTGCTAACGA 227
|||||
51 SerLeuSerValArgHisProGluTyrAsnArgLeuLeuAlaAsnAs 67
228 CCTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCATCC 277
|||||
67 PleuMetLeuIleLysLeuAspGlnSerValSerGluSerAspThrIleA 84

278 GGAGCATCAGCATGTGCTTCGCAGTGCCCTACCGCGGGGAACTCTGCCTC 327
|||||
84 rGserIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100

328 GTNTCTGGCTGGGGTCTGCTGGCGGAACGGCAGAAATGCCCTACCGTCTGCA 377
|||||

101 ValSerGlyTrpGlyLeuLeuAlaAsnGlyArgMetProThrValLeuHi 117

378 CTGCGTGAACGCTGCGGTGCTGTCTGAGGANGTCTGCAGTAAAGCTPARG 427
|||||

117 sCysValAsnValSerValValSerGlu**ValCysSerLysLeuTyrA 134

428 ACCCGCTGTACACCCACGATGTCTGCGCGCGGAGGAGCAAGACAG 477
|||||

134 sPProLeuTyrHisProSerMetPheCysAlaGlyGlyGlnAspGln 150

478 AAGACTCCTGCAACCGGTGACTCTGGGGGCGCCCTGATCTGCAACGGGTA 527
|||||

151 LysAspSerCysAsnGlyAspSerGlyGlyProLeuIleCysAsnGlyTyr 167

528 CTTGACAGGGCCTGTGTCTTTTCGGAAGGCCCGTGTGCCCACTTGCGG 577
|||||

167 rLeuGlnGlyLeuValSerPheGlyLysAlaProCysGlyGlnLeuGlyV 184

578 TGCCAGGTGTCTACACCAACTCTGCMAATTCAGTGAAGTAGAGAA 627
|||||

184 aProGlyValTyrThrAsnLeuCysLysPheThrGluTrpIleGluLys 200

628 ACCGTCACAGNCAGT 642

|||||

201 ThrValGln**Ser 205

seq_name: A_Geneseq_36:W71872

seq_documentation_block:

ID W71872 standard; Protein; 205 AA.

AC W71872;

DT 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DE13.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 127 /note- "undefined residue"

FT Misc_difference 204 /note- "undefined residue"

FT W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI: 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer

PS Example 3; Page 107-108; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour protein. The immunogen, or the DNA encoding it, can be used as a

CC vaccine for the treatment of prostate cancer. The immunogen was

CC isolated from a prostate tumour cDNA library obtained by subtracting

CC a prostate tumour cDNA expression library with a normal tissue cDNA

CC library.

CC Sequence 205 AA;

SO

alignment_scores: Quality: 1087.00 Length: 205
Ratio: 5.355 Gaps: 0

Percent Similarity: 99.024 Percent Identity: 99.024

alignment_block:

US-09-030-606-175 x W71872

US-09-030-606-175 x W71872 ..

Align seg 1/1 to: W71872 from: 1 to: 205

28 ATGGAAGAATGTTCTGCTCGGGCGCTGTCATCCGCAATGGGT 77
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVal 17
78 GCTGTACAGCCGACACTGTTCAGAACTCTACACCATCGGGCTGGGCC 127
17 LeuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
128 TGCACAGTCTTGAGGCGGACCAAGAGCCAGAGCCAGATGTTGAGGCC 177
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
178 AGCCTCTCCGTACGGCACCAGAGTACACAGACTCTTGTCTCGTTAACA 227
51 SerLeuSerValArgHisProGluTyrAsnArgLeuLeuAlaAsnAs 67
228 CCTCATGCTCATCAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCC 277
67 pleuMetLeuIleLeuLeuAspGluSerValSerGluSerAspThrIleA 84
278 GGAGCATCAGCATGCTTCGCAAGTGCCTACCGCGGGGAACTCTGCCTC 327
84 rGSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
328 GTNCTGGCTGGGCTCTGCTGGGAAACGGCAGAATGCCCTACCGTGTGCA 377
101 ValSerGlyTyrGlyLeuLeuAlaAsnGlyArgMetProThrValLeuHI 117
378 CTGCGTGAACGTGTGCGGTGCTGTGAGGANGTCTGAGTAAGCTTATG 427
117 scysValAsnValSerValSerGlu**ValCysSerLysLeuTyrA 134
428 ACCCGCTGTACACCCAGCATGTCTGCGCGCGGCGAGGCAAGACAG 477
134 spProLeuTyrHisProSerMetPheCysAlaGlyGlyGlnAspGln 150
478 AAGGACTCCTGCAACGGTGACTCTGGGGGCGCCCTGATCTGCAACGGTA 527
151 LysAspSerCysAsnGlyAspSerGlyGlyProLeuIleCysAsnGlyTyr 167
528 CTGCAAGGGCCTGTGTCTTTCGAAAGCCCGGTGTGGCAACTTGGCG 577
167 rLeuGlnGlyLeuValSerPheGlyLysAlaProCysGlyGlnLeuGly 184
578 TGCCAGGTGTCTACACCAACTCTGCAATTCAGTGTGATAGAGAAA 627
184 alProGlyValTyrThrAsnLeuCysLysPheThrGlnTrpIleGlnLys 200
628 ACCGTCAGNCCAGT 642
201 ThrValGln**Ser 205

seq_name: A_Geneseq_36:W69387

seq_documentation_block:

ID W69387 standard; Protein: 159 AA.
AC W69387;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE1 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
therapy.
OS Homo sapiens.
FH key Location/Qualifiers
FT Misc_difference 103 /note= "unspecified amino acid"
FT Misc_difference 105 /note= "unspecified amino acid"
FT Misc_difference 105 /note= "unspecified amino acid"
PN W09837418-A2.
PD 27-AUG-1998.

PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
DR N-PSDB: V58644.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 112-113; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 159 AA;

alignment_scores:
Quality: 807.00 Length: 159
Ratio: 5.206 Gaps: 0
Percent Similarity: 97.484 Percent Identity: 95.597

alignment_block:
US-09-030-606-175 x W69387 ..

Align seg 1/1 to: W69387 from: 1 to: 159

166 ATGGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGACTCTT 215
1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProle 17
216 GCTGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGT 265
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
266 CTGACACCATCCGAGCATCAGCATGCTTGCAGTGGCCCTACCGCGGG 315
34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
316 AACTCTTGCCCTGNTCTGCTGGGCTGTGCTGGGCAAGGCAAGATGCC 365
51 AsnSerCysLeuValSerGlyTyrGlyLeuLeuAlaAsnGlyArgMetPr 67
366 TACCGTCTGCACTGCGTGAACGTCGGGTGTGCTGAGGANGTCTGCA 415
67 OThrValLeuGlnCysValAsnValSerValSerGluGluValCysS 84
416 GTAAGCTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGA 465
84 erLysLeuTyrAspProLeuTyrHisProSerMetPheCysAlaGlyGly 100
466 GGGCAAGACCAAGAGACTCTGCAACGGTGACTCTGGGGGGCCCTGAT 515
101 GlyGln**Gln**AspSerCysAsnGlyAspSerGlyGlyProLeuI 117
516 CTGCAACGGGTACTTGCAAGGCGCTGTGCTTTCGAAAGCCCGGTGTG 565
117 ecysAsnGlyTyrLeuGlnGlyLeuValSerPheGlyLysAlaProCysG 134
566 GCCAAGTGGCGTCCAGAGGTGTCTACACCAACTCTGCAATTCAGTGAG 615
134 LysGlnValGlyValProGlyValTyrThrAsnLeuCysLysPheThrGlu 150
616 TGGATAGAGAAAACCGTCCAGNCCAGT 642
151 TrpIleGlnLysThrValGlnAlaSer 159

seq_name: A_Geneseq_36:w71871

seq_documentation_block:

ID W71871 standard; Protein; 159 AA.

AC W71871;

DT 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DE1.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 103 /note= "undefined residue"

FT Misc_difference 105 /note= "undefined residue"

FT Misc_difference 105 /note= "undefined residue"

PN WO9837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins -

PS used in a vaccine for the treatment of prostate cancer

PS Example 3; Page 105; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour

CC protein. The immunogen, or the DNA encoding it, can be used as a

CC vaccine for the treatment of prostate cancer. The immunogen was

CC isolated from a prostate tumour cDNA library obtained by subtracting

CC a prostate tumour cDNA expression library with a normal tissue cDNA

CC library.

SO Sequence 159 AA;

alignment_scores:

Quality: 807.00 Length: 159

Ratio: 5.206 Gaps: 0

Percent Similarity: 97.484 Percent Identity: 95.597

alignment_block:

US-09-030-606-175 x W71871 ..

Align seg 1/1 to: W71871 from: 1 to: 159

166 ATGGTGAGGCCAGCTCTCCGTACGGCAGCCAGAGTACACAGACTCTT 215

|||||

1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17

216 GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGT 265

|||||

17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34

266 CTGACACCATCCGGAGCATCAGCATTTGCTCGCAGTGCCCTACCGCGGG 315

|||||

34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50

316 AACTCTTGCCCTGCTNTCTGGCTGGGGTCTGCTGGCGAAGCGGAGATGCC 365

|||||

51 AsnSerCysLeuValSerGlyTyrGlyLeuLeuAlaAsnGlyArgMetPr 67

366 TACCGTGTGACACTGCGTGAACGTGTCGGTGTGCTGTGAGGANGTCTGCA 415

|||||

67 cThrValLeuGlnCysValAsnValSerValValSerGluGluValCysS 84

416 GTAAGCTCTATGACCCGCTGTACACACCCACAGCATGTTCTGCGCGCGGA 465

|||||

84 erLysLeuTyrAspProLeuTyrHisProSerMetPheCysAlaGlyGly 100

466 GGGCAAGACAGAGGACTCTGCAACGGTGACTCTGGGGGCGCCCTGAT 515

|||||

101 G1yGln**Gln**AspSerCysAsnGlyAspSerGlyGlyProLeuI1 117

seq_name: A_Geneseq_36:w69389

seq_documentation_block:

ID W69389 standard; Protein; 164 AA.

AC W69389;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DE14 protein.

KW Prostate tumour specific gene; human; prostate cancer; detection;

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 118 /note= "unspecified amino acid"

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-480805/41.

DR N-PSDB; V58648.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Example 1; Page 117-118; 141pp; English.

CC This sequence is encoded by a human prostate tumour specific gene, and

CC can be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC this protein sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

SO Sequence 164 AA;

alignment_scores:

Quality: 566.00 Length: 139

Ratio: 4.678 Gaps: 2

Percent Similarity: 87.050 Percent Identity: 82.014

alignment_block:

US-09-030-606-175 x W69389 ..

Align seg 1/1 to: W69389 from: 1 to: 164

28 ATGAAACGAATGTTCTGCTCGGCGCTCTGTCATCCGAGTGGGT 77

|||||

1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17

78 GCTGTACGCCGACACTGTTCCAGAACTCCTACACCATCGGGCTGGCC 127

|||||

17 lLeuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34

128 TGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCC 177

|||||

34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50

178 AGCTCTCCGTACGGCAGCCAGAGTACACAGACTCTTGCTGCTAACGA 227

CC (e.g. cancers of the bladder, prostate) or monitor HKALL regulation
CC during therapeutic intervention. Polynucleotides encoding HKALL are
CC useful to produce antisense sequences for therapeutic administration to
CC modulate/prevent HKALL expression e.g. to treat/prevent skin disorders
CC or cancer as above.
SQ Sequence 268 AA:

alignment_scores: Quality: 553.00 Length: 213
 Ratio: 3.393 Gaps: 3
Percent Similarity: 76.526 Percent Identity: 48.826

alignment_block:
us-09-030-606-175 x w94493 ..

Align seg 1/1 to: w94493 from: 1 to: 268

```
13 CAGCGGCGACTGGTCATGGAA..AACGAATTGTTCTGCGGCGTCCCT 59
|||||
56 GlnAlaIaIeuleuleuArgProAsnGlnleuTYrCysGlyAlaValle 72
60 GTGCATCCGCGAGTGGTGTCTGTACGCCGACACTGTTTCAGAACTCCT 109
|||||
72 uValHisProGlnTrpLeuLeuThrAlaAlaHisCysArgLysLysValP 89
110 ACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAAGACCAGGG 159
:::
89 heArgValArgLeuGlnHisTYrSerLeuSerProValTYrGlnSerGly 105
160 AGCCAGATGGTGAGGCCAGCCCTCTCCGTACGGCACCCAGACTACAACAG 209
:::
106 GlnGlnMetPheGlnGlnGlyValLysSerIleProHisProGlyTYrSerHi 122
210 ACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGT 259
:
122 sProGlyHisSerAsnAspLeuMetLeuIleLysLeuAsnArgArgIleA 139
260 CCGAGTCTGACACCATCCGAGACATCAGCATTGCTTCGCACTGCCCTACC 309
:::
139 rgrProThrLysAspValArgProIleAsnValSerSerHisCysProSer 155
310 GCGGGGAACCTCTGCCTCTGCTNTGCTGGGCTGCTGTCGCGCAACGGCAG 359
|||||
156 AlaGlyThrLysCysLeuValSerGlyTrpGlyThrThrLysSerProG1 172
360 AATG.....CCTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
:::
172 nValHisPheProLysValLeuGlnCysLeuAsnIleSerValLeuSerG 189
404 AGGANGTCTGAGTAAGCTTATGACCCGCTGTACACCCAGCATGTTC 453
:::
189 InLysArgCysGlnAspAlaTYrProArgGlnIleAspAspThrMetPhe 205
454 TGCGCCGCGGAGGGCAAGACAGAGCAAGACTCCTGCAACGGTGACTCTGG 503
|||||
206 CysAla...GlyAspLysAlaGlyArgAspSerCysGlnGlyAspSerG1 221
504 GGGGCCCTGATCTGCAACGGGTACTTGCAAGGCTGTGTCTTTGCGAA 553
|||||
221 yGlyProValValCysAsnGlySerLeuGlnGlyLeuValSerTrpGlyA 238
554 AAGCCCGGTGTGGCCAACTTGCGGTGCCAGGTGTCTACACCAACTCTGC 603
|||||
238 sPTyrProCysAlaArgProAsnArgProGlyValTYrThrAsnLeuCys 254
604 AAATTCACGTGAGTGATAGAGAAACCGTCCAGNCACAGT 642
|||||
255 LysPheThrLysTrpIleGlnGlnThrIleGlnAlaAsn 267
```

seq_name: A_Geneseq_36:R67888

seq_documentation_block:

ID R67888 standard; Protein; 253 AA.
AC R67888;
DT 09-AUG-1995 (first entry)
DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; Ichthyoses; eczema.
OS Homo sapiens.
PN W09500651-A.
PD 05-JAN-1995.
PF 20-JUN-1994; IB0166.
PR 18-JUN-1993; DK-000725.
PA (SYMB-) SYMBICOM AB.
PI Egelrud T, Hansson L;
DR WPI; 95-052088/07.
DR N-PSDB; Q81203.
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides,
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
PS Disclosure; Page 97; 137pp; English.
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), Ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
SQ Sequence 253 AA:

alignment_scores: Quality: 530.50 Length: 214
 Ratio: 3.255 Gaps: 3
Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:

us-09-030-606-175 x R67888 ..

Align seg 1/1 to: R67888 from: 1 to: 253

```
1 GCGGACCCCTGGCAGGCGGCGACTGTCATGGAACGAATTGTTCTGCTC 50
:::
40 SerHisProTrpGlnValAlaIeuleuSerGlyAsnGlnleuHisCysG1 56
51 GGGGCTCCTGTGTCATCCGCGAGTGGGTGCTGTACGCCGACACTGTTCC 100
|||||
56 yGlyValleuValAsnGlnArgTrpValleuThrAlaAlaHisCysLysM 73
101 AGAATTCCTACACCATCGGCGCTGGGCTGCACAGTCTTGAGCCGACCAA 150
:::
73 eTAsnGlnTYrThrValHisLeuGlySerAspThrleu..GlyAspArg 88
151 GAGCCAGGGAGCCAGATGTGGAGGCCAGCCCTCTCCGTACGGCACCCAGA 200
:::
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG1 103
201 GTACACAGACTCTTGCTGCTAACGACCTCATGCTCATCAAGTTGACG 250
|||||
103 TYrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsnS 120
251 AATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCATTGCTTCGAG 300
:::
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
301 TGCCCTACCCGGGGGAGACTCTTGCCCTGCTNTGCTGGGCTGCTGCGC 350
|||||
137 CysGlnProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153
351 GAACGGC.....AGATGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCG 394
:::
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170
395 TGGTGTCTGAGGANGTCTGAGTAAGCTTATGACCCGCTGTACACACCC 444
:::
|||
```


alignment_scores:

Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

alignment_block:

US-09-030-606-175 x W10694 ..

Align seg 1/1 to: W10694 from: 1 to: 260

```
1 GCGCAGCCCTGGCAGCGCGCACTGGTCATGGAACGAATTGTTCTGCTC 50
  :::::::::::::::::::::: :: ::::::::::::::
43 SerGlnProTPrGlnAlaAlaLeuPheGlnGlyAlaArgLeuIleCysG1 59
51 GGGCGTCTGTGGCATCCGCAGTGGGTGTGTACGCCGACACTGTTTC 100
  :::::::::::::: :::::::::::::::
59 YGlyValLeuValGlyAspArgTPrValLeuThrAlaAlaHisCysLysL 76
101 AGAACTCCTAACCATCGGGCTGGCTGCACAGTCTTGAGCCGACCAA 150
  ::::::::::::::: ||||| ||||| ::::
76 YsGlnLysTyrSerValArgLeuGlyAspHisSerLeuGlnSerArgAsp 92
151 GAGCCAGGAGCCAGATGTGTGAGGCCACCTCTCCGTACGGCACCAGA 200
  ::||| ::||| ::::::::::: ||::: |||||
93 GlnPro..GluGlnGluIleGlnValAlaGlnSerIleGlnHisProCy 108
201 GTACACAGACTCTTG.....CTGCCTAACGACCTCATGCTCATCA 241
  |||||::: :::::||:::|||||:::
108 sTyraSnaSerAsnProGluAspHisSerHisAspIleMetLeuIleA 125
242 AGTTGACGAATCCGTGTCGAGTCTGACACCATCCGGAGCATCAGCAT 291
  ::||::: ||::: ||::: ::||| :::::
125 rGleuGlnAsnSerAlaAsnLeuGlyAspLysValLysProValGlnLeu 141
292 GCTTCGAGTGCCTACCGGGGGAAGTCTGCTGCTGCTGCTGCTGGG 341
  ||::: ||||| ::|||::: ||::: ||||| |||||
142 AlaAsnLeuCysProLysValGlyGlnLysCysIleIleSerGlyTPrG1 158
342 TCTGCTGGCAGCGCAGAG.....ATGCCTAACGCTGCTGACCTGCGTGA 385
  | ::::: :::: ::|||::: ||::: ||:::
158 yThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysAlaG 175
386 ACGTGTGGTGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTG 435
  ::||::: ||::: ||::: ||::: ||::: ||
175 luValLysIleTyrSerGlnAsnLysCysGluArgAlaTyrProGlyLys 191
436 TACCACCCAGCATGTCTGCGCGCGGCGGAGGCAAGACAGAGACTC 485
  ::||| ::||| ||||| ::::: ||::: ||:::
192 IleThrGluGlyMetValCysAla..GlySerSerAsnGlyAlaAspTh 207
486 CTGCAACGGTACTGTGGGGGCCCTGATCTGCAACGGGTACTTGACAG 535
  ::||::: ||||| ||||| ||||| ::||| |||||
207 rCysGlnGlyAspSerGlyGlyProLeuValCysaspGlyMetLeuGlnG 224
536 GCCTGTGTCTTTTCGAAAAGCCCGTGTGGCCAACCTTGCGCTGAGGT 585
  ||::: ||::: ||::: ||||| ::||| |||||
224 yIleThrSerTPrGlySerAspProCysGlyLysProGluLysProGly 240
586 GTCTACACCAACCTCTGCAATATCAGTGGATAGAGAAACCGTC 633
  ||||| ||||| ::|||::: ||||| ::|||:::
241 ValTyrThrLysIleCysArgTyrThrThrIleLysLysThrMet 256
```

seq_name: A_Geneseq_36:W12393

seq_documentation_block:

ID W12393 standard; Protein; 260 AA.

AC W12393;

DT 15-MAY-1997 (first entry)

DE Mouse neuropsin protein.

KW Mouse; neuropsin; hippocampus; lambda gt10; primer; PCR; amplification;

KW polymerase chain reaction; serine protease domain; nerve growth factor;

KW NGF; insect cell; virus; expression vector; transfection;

KW cerebral disease.

OS Mus musculus.

PN J08311099-A.

PD 26-NOV-1996.

PF 13-MAR-1996; 056367.

PR 14-MAR-1995; JP-054584.

PA (SHIO/) SHIOZAKA S.

DR WPI; 97-061812/06.

DR N-PSDB; T63251.

PT Nucleic acid encoding neuropsin - for producing neuropsin, useful

PR for diagnosis and treatment of cerebral disease

PS Claim 1; Page 6-7; 9pp; Japanese.

CC This is the amino acid sequence of a novel mouse protein designated

CC neuropsin. The encoding gene was isolated from a mouse hippocampal

CC cDNA library in lambda gt10 using a cloned, amplified fragment of

CC the gene (clone B41; T63254). This fragment was amplified using

CC primers T63252-3. The primers were synthesised based on the serine

CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated

CC 6 positive clones, of which clone NP5 contained the longest insert

CC (this sequence). The protein has a molecular weight of around 26 kD.

CC It has 43% homology with EGF-BP, 41 % with NGF-gamma; 39% with NGF-alpha;

CC 38% with trypsin and 18% with tPA. The protein can be used for clinical

CC diagnosis and treatment of cerebral diseases.

Sequence 260 AA;

alignment_scores:

Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

alignment_block:

US-09-030-606-175 x W12393 ..

Align seg 1/1 to: W12393 from: 1 to: 260

```
1 GCGCAGCCCTGGCAGCGCGCACTGGTCATGGAACGAATTGTTCTGCTC 50
  :::::::::::::::::::::: :: ::::::::::::::
43 SerGlnProTPrGlnAlaAlaLeuPheGlnGlyAlaArgLeuIleCysG1 59
51 GGGCGTCTGTGGCATCCGCAGTGGGTGTGTACGCCGACACTGTTTC 100
  :::::::::::::: :::::::::::::::
59 YGlyValLeuValGlyAspArgTPrValLeuThrAlaAlaHisCysLysL 76
101 AGAACTCCTAACCATCGGGCTGGCTGCACAGTCTTGAGCCGACCAA 150
  ::::::::::::::: ||||| ||||| ::::
76 YsGlnLysTyrSerValArgLeuGlyAspHisSerLeuGlnSerArgAsp 92
151 GAGCCAGGAGCCAGATGTGTGAGGCCACGCTCTCCGTACGGCACCAGA 200
  ::||| ::||| ::::::::::: ||::: |||||
93 GlnPro..GluGlnGluIleGlnValAlaGlnSerIleGlnHisProCy 108
201 GTACACAGACTCTTG.....CTGCCTAACGACCTCATGCTCATCA 241
  |||||::: :::::||:::|||||:::
108 sTyraSnaSerAsnProGluAspHisSerHisAspIleMetLeuIleA 125
242 AGTTGACGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCAT 291
  ::||::: ||::: ||::: ::||| :::::
125 rGleuGlnAsnSerAlaAsnLeuGlyAspLysValLysProValGlnLeu 141
292 GCTTCGAGTGCCTACCGGGGGAAGTCTGCTGCTGCTGCTGCTGGG 341
  ||::: ||||| ::|||::: ||::: ||||| |||||
142 AlaAsnLeuCysProLysValGlyGlnLysCysIleIleSerGlyTPrG1 158
342 TCTGCTGGCAGCGCAGAG.....ATGCCTAACGCTGCTGACCTGCGTGA 385
  | ::::: :::: ::|||::: ||::: ||:::
158 yThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysAlaG 175
386 ACGTGTGGTGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTG 435
  ::||::: ||::: ||::: ||::: ||::: ||
175 luValLysIleTyrSerGlnAsnLysCysGluArgAlaTyrProGlyLys 191
436 TACCACCCAGCATGTCTGCGCGCGGAGGAGGCAAGACAGAGACTC 485
  ::||| ::||| ||||| ::::: ||::: ||:::
```


Align seg 1/1 to: W22985 from: 1 to: 244

```
1 GCGCAGCCCTGGCAGGGCGGCACTGGTCATGGAAAAAGCAATTGTCTGCTC 50
   ::::::::::::::::::::
32 SerHisProTyrGlnAlaAlaLeuTyrThrSerGlnHisLeuLeuCysG1 48
   ::::::::::::::::::::
51 GGGCGTCCGTGGTCATCCGCAGTGGGTGCTGCAGCCGACACTGTTC 100
   ::::::::::::::::::::
48 yG1yValLeuIleHisProLeuTyrPValLeuThrAlaAlaHisCysLysL 65
   ::::::::::::::::::::
101 AGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTT.....GAG 141
   ::::::::::::::::::::
65 ySProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu 81
   ::::::::::::::::::::
142 GCCGACCAAGAGCCAGGAGCCAGATGGTGAAGCCAGCCCTTCGGTACG 191
   ::::::::::::::::::::
82 SerSerGlnGlnGln...SerSerValValArgAla.....ValI1 94
   ::::::::::::::::::::
192 GCACCCAGAGTACACAGACTCTTGTGCTGCTAAGACCTCATGCTCATCA 241
   ::::::::::::::::::::
94 eHisProAspTyrAspAlaAlaSerHisAspGlnAspIleMetLeuLeuA 111
   ::::::::::::::::::::
242 AGTTGACGAATCCGTGCCAGTCTGCACACCATCCGAGCATGACCATT 291
   ::::::::::::::::::::
111 rGleuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127
   ::::::::::::::::::::
292 GCTTCGACGTGCCCTACCGCGGGGGAACCTTGCCCTGTCGCTGGGG 341
   ::::::::::::::::::::
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTyrP1 144
   ::::::::::::::::::::
342 TCTGCTGGCGAAGCGCAGAATGCCCTACCGTGCCTGCACCTGCGTGAACGT 391
   ::::::::::::::::::::
144 yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTyrIleH 161
   ::::::::::::::::::::
392 CGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC 441
   ::::::::::::::::::::
161 lSLeuValSerArgGluGluCysGlnHisAlaTyrProGlyGlnIleThr 177
   ::::::::::::::::::::
442 CCCAGCATGTTCTGCGCGCGGCGAGGCGCAAGACAGAGACTCTGCA 491
   ::::::::::::::::::::
178 GlnAsnMetLeuCysAlaGlyAspGlnLysTyrGlyLysAspSerCysG1 194
   ::::::::::::::::::::
492 CGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGACAGGCTTG 541
   ::::::::::::::::::::
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeuV 211
   ::::::::::::::::::::
542 TGTCTTTCGAAAGCCCCGTGTGCCCACTTGCGGTGCCAGGTGTCTAC 591
   ::::::::::::::::::::
211 aISerTyrGlyAsnIleProCysGlySerLysGluLysProGlyValTyr 227
   ::::::::::::::::::::
592 ACCAACCCTCTGCAAAATTCAGTGAATGAGAAAAACCGTCCAG 636
   ::::::::::::::::::::
228 ThrAsnValCysArgTyrThrAsnTyrIleGlnLysThrIleGln 242
```


This Page Blank (uspto)

|||||
Db 361 ATGCCTACCGTCTGCACCTGCGTGAACGTGTGCGTGTCTGAGGANGTCTGCAGTAAG 420
QY 421 CTCTATGACCCGCTGTACCACCCCAAGCATGTCTGCGCCGGGAGGCGCAAGACAGAAAG 480
Db 421 CTCTATGACCCGCTGTACCACCCCAAGCATGTCTGCGCCGGGAGGCGCAAGACAGAAAG 480
QY 481 GACTCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAAGGCTT 540
Db 481 GACTCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAAGGCTT 540
QY 541 GTGTCTTTGGAAGCCCGTGTGGCCAACTTGCGCTGCCAGGTGTCTACACCAACCTC 600
Db 541 GTGTCTTTGGAAGCCCGTGTGGCCAACTTGCGCTGCCAGGTGTCTACACCAACCTC 600
QY 601 TGCATAATTCAGTGAATAGAGAAACCGTCCAGNCCAGTTAACTCTGGGGAGCTGGA 660
Db 601 TGCATAATTCAGTGAATAGAGAAACCGTCCAGNCCAGTTAACTCTGGGGAGCTGGA 660
QY 661 ACCCATGAATTCAGCCCAATACATCTCTGCGGAANGAATTCAGGAATATCTGTCCCA 720
Db 661 ACCCATGAATTCAGCCCAATACATCTCTGCGGAANGAATTCAGGAATATCTGTCCCA 720
QY 721 GCGCCTCTCTCCCTCAGGCGCCAGGAGTCCAGGCGCCCTCTCTCTCAAAACCAAG 780
Db 721 GCGCCTCTCTCCCTCAGGCGCCAGGAGTCCAGGCGCCCTCTCTCTCAAAACCAAG 780
QY 781 GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCCTCNT 840
Db 781 GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCCTCNT 840
QY 841 CCNTCAGACCCAGAGTCCAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCAGC 900
Db 841 CCNTCAGACCCAGAGTCCAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCAGC 900
QY 901 CCNTCCTCCGTCAGACCCAGGCGGTGCAGGCGCCCAACCCCTCNTCCNTCAGAGTCAGAG 960
Db 901 CCNTCCTCCGTCAGACCCAGGCGGTGCAGGCGCCCAACCCCTCNTCCNTCAGAGTCAGAG 960
QY 961 TCAGAGCCCAACCCCT 1020
Db 961 TCAGAGCCCAACCCCT 1020
QY 1021 TCAGAGCCCAAGCGGTCCCAATGCCAAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 1080
Db 1021 TCAGAGCCCAAGCGGTCCCAATGCCAAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 1080
QY 1081 NGTTCAGCCCAACCTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 NGTTCAGCCCAACCTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
QY 1141 ATAAAGTNTAAGAGAGCGCAAAAAA 1167
Db 1141 ATAAAGTNTAAGAGAGCGCAAAAAA 1167

RESULT 2
V61252
ID V61252 standard; cDNA: 1167 BP.
AC V61252;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.

DR P-PSDB; W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 98.6%; Score 1151; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.7e-278;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCAGCCCTGGCAGCGCGGCACTGGTCAATGGAAGCAATTTGTTCTCGGGCGTCTG 60
Db 1 GCGCAGCCCTGGCAGCGCGGCACTGGTCAATGGAAGCAATTTGTTCTCGGGCGTCTG 60
QY 61 GTGCATCCGAGTGGGTGCTGTACGCGCCGACACTGTTCCAGAACTCTACACCATGCGG 120
Db 61 GTGCATCCGAGTGGGTGCTGTACGCGCCGACACTGTTCCAGAACTCTACACCATGCGG 120
QY 121 CTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCAGATGTGAGGCGCAGC 180
Db 121 CTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCAGATGTGAGGCGCAGC 180
QY 181 CTCTCCGTACGGCAGCCAGAGTACAAACAGCTTGTCTGCTTAACGACCTCATGCTCATC 240
Db 181 CTCTCCGTACGGCAGCCAGAGTACAAACAGCTTGTCTGCTTAACGACCTCATGCTCATC 240
QY 241 AAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGATCAGCATTGCTTCGAG 300
Db 241 AAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGATCAGCATTGCTTCGAG 300
QY 301 TGCCCTACCGCGGGGAACTCTTGCCCTGTTCTGGCTGGGGTCTCTGCGAAGCGGAGA 360
Db 301 TGCCCTACCGCGGGGAACTCTTGCCCTGTTCTGGCTGGGGTCTCTGCGAAGCGGAGA 360
QY 361 ATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 361 ATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 421 CTCTATGACCCGCTGTACCACCCCAAGCATGTCTGCGCGGCGGAGGCGCAAGACAGAG 480
Db 421 CTCTATGACCCGCTGTACCACCCCAAGCATGTCTGCGCGGCGGAGGCGCAAGACAGAG 480
QY 481 GACTCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAAGGCTT 540
Db 481 GACTCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAAGGCTT 540
QY 541 GTGTCTTTGGAAGCCCGTGTGGCCAACTTGCGCTGCCAGGTGTCTACACCAACCTC 600
Db 541 GTGTCTTTGGAAGCCCGTGTGGCCAACTTGCGCTGCCAGGTGTCTACACCAACCTC 600
QY 601 TGCATAATTCAGTGAATAGAGAAACCGTCCAGNCCAGTTAACTCTGGGGAGCTGGA 660
Db 601 TGCATAATTCAGTGAATAGAGAAACCGTCCAGNCCAGTTAACTCTGGGGAGCTGGA 660
QY 661 ACCCATGAATTCAGCCCAATACATCTCTGCGGAANGAATTCAGGAATATCTGTCCCA 720
Db 661 ACCCATGAATTCAGCCCAATACATCTCTGCGGAANGAATTCAGGAATATCTGTCCCA 720
QY 721 GCGCCTCTCTCCCTCAGGCGCCAGGAGTCCAGGCGCCCTCTCTCTCAAAACCAAG 780
Db 721 GCGCCTCTCTCCCTCAGGCGCCAGGAGTCCAGGCGCCCTCTCTCTCAAAACCAAG 780
QY 781 GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCCTCNT 840
Db 781 GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCCTCNT 840

| | | | |
|----|------|---|------|
| OY | 841 | CCNTCAGACCCGAGAGTCCAGCCCCCTCTTCNTCAGACGCAGAGTCCAGACCCCACC | 900 |
| Db | 841 | CCNTCAGACCCGAGAGTCCAGCCCCCTCTTCNTCAGACGCAGAGTCCAGACCCCACC | 900 |
| OY | 901 | CCNTCNTCCGTCAAGACCCAGGGGTGACGCCCCCAACCCCTCMTCCNTCAGAGTCAAGG | 960 |
| Db | 901 | CCNTCNTCCGTCAAGACCCAGGGGTGACGCCCCCAACCCCTCMTCCNTCAGAGTCAAGG | 960 |
| OY | 961 | TCCAAGCCCCCAACCCCTCGTTCCCCACAGACCAGAGTNCAGGTCCAGCCCCCTCTCCC | 102 |
| Db | 961 | TCCAAGCCCCCAACCCCTCGTTCCCCACAGACCAGAGTNCAGGTCCAGCCCCCTCTCCC | 102 |
| OY | 1021 | TCAGACCCAGCGGTCCAATGCCACCTAGANTNNTCCCTGTACAGATGCCCCCTGTGCA | 108 |
| Db | 1021 | TCAGACCCAGCGGTCCAATGCCACCTAGANTNNTCCCTGTACAGATGCCCCCTGTGCA | 108 |
| OY | 1081 | NGTGACCCCAACCTTACCAGTGTGTTTTTCATTTTTTGTCCCTTTCCCTAGATCCAGA | 114 |
| Db | 1081 | NGTGACCCCAACCTTACCAGTGTGTTTTTCATTTTTTGTCCCTTTCCCTAGATCCAGA | 114 |
| OY | 1141 | ATAAAGTNTAAGAGAAGCGCAAAAAAA | 1167 |
| Db | 1141 | ATAAAGTNTAAGAGAAGCGCAAAAAAA | 1167 |

```

RESULT      3
V58644
ID      V58644 standard; cDNA; 1248 BP.
AC      V58644;
DT      08-DEC-1998 (first entry)
DE      Prostate tumour specific gene clone DE1.
KW      Prostate tumour specific gene; human; prostate cancer; detection;
        therapy; ss.
KW      Homo sapiens.
FH      Key
FT      CDS
FT      Location/Qualifiers
        217..696
        /*tag= a
PN      WO9837418-A2.
PD      27-AUG-1998.
PF      25-FEB-1998; U03690.
PR      09-FEB-1998; US-904809.
PR      25-FEB-1997; US-806596.
PR      01-AUG-1997; US-904809.
PA      (CORI-) CORIXA CORP.
PI      Dillon DC, Xu J;
DR      WPI: 98-480805/41.
DR      P-PSDB: W69387.
PT      Novel human prostate specific tumour protein and fragments - useful
PT      for detecting and treating prostate cancers
PS      Claim 1; Page 112; 141pp; English.
CC      This sequence represents a human prostate tumour specific gene, and can
CC      be used in the method of the invention. The method is for detecting
CC      prostate cancer comprises contacting a biological sample with an agent
CC      able to bind an immunogenic portion of a prostate protein (such as
CC      encoded by this sequence). An antibody which binds to an immunogenic
CC      portion of the prostate protein, and the method can be used to detect,
CC      monitor progression of, or treat prostate cancers. The antibody may
CC      also be conjugated to a therapeutic agent for use in therapy of prostate
CC      cancers.
SQ      Sequence      1248 BP;      288 A;      424 C;      303 G;      228 T;

Query Match
Best Local Similarity 95.2%; Score 1111.2; DB 1; Length 1248;
Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps

QY      2 CGCAGCCCTGGCAGCGCGGCACTGTCATGGAACGAATTTCTGCTCGGGCGTCTGG 61
        |||||||
DB      41 CGCAGCCCTGGCAGCGCGGCACTGTCATGGAACGAATTTCTGCTCGGGCGTCTGG 100

QY      62 TGCATCCGCAGTGGGTGCTGTCAAGCCGCACACACTGTTCCAGAA-----CTCCT 109
        |||||||
DB      101 TGCATCCGCAGTGGGTGCTGTCAAGCCGCACACACTGTTCCAGAAAGTGAGTGCAGAGCTCCT 160

```

| | | | |
|----|------|---|------|
| QY | 110 | ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG | 169 |
| Db | 161 | ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG | 220 |
| QY | 170 | TGAGAGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCGCTAACGACC | 229 |
| Db | 221 | TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCGCTAACGACC | 280 |
| QY | 230 | TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGGAGCATCAGCA | 289 |
| Db | 281 | TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGGAGCATCAGCA | 340 |
| QY | 290 | TTGCTTGCAGTGGCCCTACCGCGGGGAACCTTTGCCCTCGTNTCTGGCTGGGGTCTGTGG | 349 |
| Db | 341 | TTGCTTGCAGTGGCCCTACCGCGGGGAACCTTTGCCCTCGTNTCTGGCTGGGGTCTGTGG | 400 |
| QY | 350 | CGAAGCGCAGAAATGCGCTTACCGTGTGCACTGCGTGAACGNTGTGGTGTCTGAGGANG | 409 |
| Db | 401 | CGAAGCGCAGAAATGCGCTTACCGTGTGCACTGCGTGAACGNTGTGGTGTCTGAGGANG | 460 |
| QY | 410 | TTCTGCAGTAAGCTCTATGACCCCGCTGTACCACCCAGACATGTCTCGCGCCGCGAGGGC | 469 |
| Db | 461 | TTCTGCAGTAAGCTCTATGACCCCGCTGTACCACCCAGACATGTCTCGCGCCGCGAGGGC | 520 |
| QY | 470 | AAGAACCGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCCGTATCTGCAACGGGACT | 529 |
| Db | 521 | AAGAACCGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCCGTATCTGCAACGGGACT | 580 |
| QY | 530 | TGCAGGGCCTGTGTCTTTCGGAAGAACCCCGTGTGGCCAATTGGCGTGCCAGGTGCT | 589 |
| Db | 581 | TGCAGGGCCTGTGTCTTTCGGAAGAACCCCGTGTGGCCAATTGGCGTGCCAGGTGCT | 640 |
| QY | 590 | ACACCAACCTCTGCAAAATTCAGTGATAGAGAAACCGTCCAGNCCAGTTAATCT | 649 |
| Db | 641 | ACACCAACCTCTGCAAAATTCAGTGATAGAGAAACCGTCCAGNCCAGTTAATCT | 700 |
| QY | 650 | GGGAGCTGGGAACCCATGAAATTGACCCCCCAATACATCTCGGGAANGAATTCAGGAAT | 709 |
| Db | 701 | GGGAGCTGGGAACCCATGAAATTGACCCCCCAATACATCTCGGGAANGAATTCAGGAAT | 760 |
| QY | 710 | ATCTGTCCAGCCCCCTCTCTCCCTCAGGCCCAGGAGTCCAGGGCCCCCAGCCCCCTCTCC | 769 |
| Db | 761 | ATCTGTCCAGCCCCCTCTCTCCCTCAGGCCCAGGAGTCCAGGGCCCCCAGCCCCCTCTCC | 820 |
| QY | 770 | TCAAAACCAAGGATACAGATCCCCAGCCCCCTCTCTCCCTCAGACCCAGAGTCCAGACCCC | 829 |
| Db | 821 | TCAAAACCAAGGATACAGATCCCCAGCCCCCTCTCTCTCCCTCAGACCCAGAGTCCAGACCCC | 880 |
| QY | 830 | CAGCCCCCTCTCTCCCTCAGACCCAGGAGTCCAGCCCCCTCTCTCTCCCTCAGACCCAGGATCCA | 889 |
| Db | 881 | CAGCCCCCTCTCTCCCTCAGACCCAGGAGTCCAGCCCCCTCTCTCTCTCCCTCAGACCCAGGATCCA | 940 |
| QY | 890 | GACCCCCCAGCCCNCTCCTCCGTAGACCCAGGGGTGAGGGCCCCCAACCCCTCCTCCTC | 949 |
| Db | 941 | GACCCCCCAGCCCNCTCCTCCGTAGACCCAGGGGTGAGGGCCCCCAACCCCTCCTCCTC | 1000 |
| QY | 950 | AGAGTCAGAGGTCCAAGCCCCCAACCCCTCGTTCCCGCAGACCAGAGGTNCAAGTCCAG | 1009 |
| Db | 1001 | AGAGTCAGAGGTCCAAGCCCCCAACCCCTCATTTCCCGCAGACCAGAGTCCAGGTCCAG | 1060 |
| QY | 1010 | CCCTCTCTCCCTCAGACCCAGCGGTCCAATGCCACTAGANTNTCCCTGTACACAGTGCC | 1069 |
| Db | 1061 | CCCTCTCTCCCTCAGACCCAGCGGTCCAATGCCACTAGANTNTCCCTGTACACAGTGCC | 1120 |
| QY | 1070 | CCCTTGTGGCANGTTGACCCCAACCTTACCAGTTGGTTTTCATTTTGTCCCTTCCCC | 1129 |
| Db | 1121 | CCCTTGTGGCANGTTGACCCCAACCTTACCAGTTGGTTTTCATTTTGTCCCTTCCCC | 1180 |
| QY | 1130 | TAGATCCAGAAATAAAGTNTAAGAGAAGCGCAAAAAA | 1167 |
| Db | 1181 | TAGATCCAGAAATAAAGTNTAAGAGAAGNGCAAAAAA | 1218 |

| QY | 1130 | TAGATCCAGAAATAAGTNTTAAGACAGAGCGCAAAAAA | 1167 |
|----|------|--|------|
| | | | |
| Db | 1181 | TAGATCCAGAAATAAGTTTAAGACAGAGNGCAAAAAAA | 1218 |

RESULT 4
ID V61249 standard; cDNA: 1248 BP.
AC V61249;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE1.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
DR P-PSDB: W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 95.2%; Score 1111.2; DB 1; Length 1248;
Best Local Similarity 96.7%; Pred. No. 2.3e-268;
Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps 1;

QY 2 CGCAGCCCTGGCAGGGGCACTGTGCATGAAACGAATGTTCTGCTGGGCGTCTCG 61
Db 41 CGCAGCCCTGGCAGGGGCACTGTGCATGAAACGAATGTTCTGCTGGGCGTCTCG 100
QY 62 TGCATCCGCAAGTGGGTGCTGTACAGCCGCACTGTTCCAGAA-----CTCCT 109
Db 101 TGCATCCGCAAGTGGGTGCTGTACAGCCGCACTGTTCCAGAAAGTGTGAGAGCTCT 160
QY 110 ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATG 169
Db 161 ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATG 220
QY 170 TGGAGGCCAGCTCTCCGTACGGCACCAGATACAAAGACTCTGCTCGCTAACGACC 229
Db 221 TGGAGGCCAGCTCTCCGTACGGCACCAGATACAAAGACTCTGCTCGCTAACGACC 280
QY 230 TCATGCTCATCAAGTTGACAGAAATCCGTGCCAGTCTGACACCAATCCGGAGATCAGCA 289
Db 281 TCATGCTCATCAAGTTGACAGAAATCCGTGCCAGTCTGACACCAATCCGGAGATCAGCA 340
QY 290 TTGCTTCGCAAGTGGCTTACCGGGGGAACCTTTCCTCGTNTCTGGCTGGGCTCTGCTGG 349
Db 341 TTGCTTCGCAAGTGGCTTACCGGGGGAACCTTTCCTCGTNTCTGGCTGGGCTCTGCTGG 400
QY 350 CGAACGGCAGAAATGCTTACCGGTGCTGACACTGCTGAAGCTGTGGTGGTGTCTGAGGANG 409
Db 401 CGAACGGCAGAAATGCTTACCGGTGCTGACACTGCTGAAGCTGTGGTGGTGTCTGAGGANG 460
QY 410 TCTGAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGCGAGAGGC 469
Db 461 TCTGAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGCGAGAGGC 520
QY 470 AAGACCAAGAGACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT 529
Db 521 AAGACCAAGAGACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT 580
QY 530 TGCAGGGCCTGTGTCTTTGGAAGAGCCCGGTGGCCAACTTGCGCTGCCAGGTGTCT 589

Db 581 TGCAGGGCCTGTGTCTTTTCGAAAGAGCCCGTGTGGCCAAAGTTGGCGTCCAGGTGTCT 640
QY 590 ACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAACCGTCCAGNCCAGTTACTCT 649
Db 641 ACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAACCGTCCAGNCCAGTTACTCT 700
QY 650 GGGACTGGGAACCCATGAATATGACCCCAATATCATCTCTGGGAANGAATTCAGGAAT 709
Db 701 GGGACTGGGAACCCATGAATATGACCCCAATATCATCTCTGGGAANGAATTCAGGAAT 760
QY 710 ATCTGTCCAGCCCTCTCTCTCAGGCCCCAGAGTCCAGGCCCCAGCCCTCTCTCC 769
Db 761 ATCTGTCCAGCCCTCTCTCTCAGGCCCCAGAGTCCAGGCCCCAGCCCTCTCTCC 820
QY 770 TCAAACCAAGGTTACAGATCCCGAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCC 829
Db 821 TCAAACCAAGGTTACAGATCCCGAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCC 880
QY 830 CAGCCCTCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCTCAGACCCAGAGTCC 889
Db 881 CAGCCCTCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCTCAGACCCAGAGTCC 940
QY 890 GACCCCT 949
Db 941 GACCCCT 1000
QY 950 AGAGTCAAGGTTCCAAAGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1009
Db 1001 AGAGTCAAGGTTCCAAAGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1060
QY 1010 CCCCT 1069
Db 1061 CCCCT 1120
QY 1070 CCCTGTGGCAGTTGACCCCAACCTTACAGTGTGTTTTCATTTTGTCTCTCTCTCT 1129
Db 1121 CCCTGTGGCAGTTGACCCCAACCTTACAGTGTGTTTTCATTTTGTCTCTCTCTCT 1180
QY 1130 TAGATCCAGAAATAAGTNTAAGAGAGCGCAAAAAA 1167
Db 1181 TAGATCCAGAAATAAGTNTAAGAGAGNGCAAAAAA 1218

RESULT 5
ID V58645 standard; cDNA: 1265 BP.
AC V58645;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
PT Novel human prostate specific protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.

| | | | | | | |
|----|-----------------------|---------------------|---------------------|-----------------------|---------------|---------|
| SQ | Sequence | 1265 BP; | 256 A; | 432 C; | 321 G; | 245 T; |
| | Query Match | 88.8%; | Score 1036.2; | DB 1; | Length 1265; | |
| | Best Local Similarity | 91.0%; | Pred. No. 1.2e-249; | | | |
| | Matches 1137; | Conservative | 0; | Mismatches 29; | Indels 83; | Gaps 1; |
| QY | 2 | CGCAGCCCTGGCAGCGGGC | ACTGGTCATGAAAACGAAT | GTGCTGCGGGCTCCTGG | 61 | |
| DB | 14 | CGCAGCCCTGGCAGCGGGC | ACTGGTCATGAAAACGAAT | GTGCTGCGGGCTCCTGG | 73 | |
| QY | 62 | TGCATCCGACGTGGGTGCT | GTCAGCCGACACTGTTCC | AGAACTCCTACACCATCGGGC | 121 | |
| DB | 74 | TGCATCCGACGTGGGTGCT | GTCAGCCGACACTGTTCC | AGAACTCCTACACCATCGGGC | 133 | |
| QY | 122 | TGGGCTGCACAGTCTTGAG | CGCCGACCAAGACCAGGAG | CCAGATGGTGAGGCCAGCC | 181 | |
| DB | 134 | TGGGCTGCACAGTCTTGAG | CGCCGACCAAGACCAGGAG | CCAGATGGTGAGGCCAGCC | 193 | |
| QY | 182 | TCTCCGTACGGCACCCAG | AGATACACAGACTCTGCT | CGCTAACGACCTCATGCTCA | 241 | |
| DB | 194 | TCTCCGTACGGCACCCAG | AGATACACAGACTCTGCT | CGCTAACGACCTCATGCTCA | 253 | |
| QY | 242 | AGTTGACGAATCCGTGTCC | GAGTCTGACACCATCCGG | AGCATCAGATTGCTTCGCA | 301 | |
| DB | 254 | AGTTGACGAATCCGTGTCC | GAGTCTGACACCATCCGG | AGCATCAGATTGCTTCGCA | 313 | |
| QY | 302 | GCCCTACCGCGGGGAAC | TCTTGCCCTGTTCTGGGT | GTGCTGCGGAAC----- | 355 | |
| DB | 314 | GCCCTACCGCGGGGAAC | TCTTGCCCTGTTCTGGGT | GTGCTGCGGAACGGGTGAGC | 373 | |
| QY | 355 | ----- | ----- | ----- | 355 | |
| DB | 374 | TCACGGGTGTGTCTGCCC | CTTCAAGAGAGTCTCTG | CCCAATCGCGGGGTGACCC | 433 | |
| QY | 355 | ----- | GGCAGATGCTTACCGGT | GCTGCACACTGCGTGAAC | GTGCGTGT 398 | |
| DB | 434 | AGAGCTGTGCTCCAGGCA | AGATGCTTACCGGTGCT | GCAAGTGCATGCGTGTGT | 493 | |
| QY | 399 | GTCAGGANGTCTGACGTA | AGCTCTATGACCCGCTGT | ACACCCAGCATGTTCTGCGC | 458 | |
| DB | 494 | GTCAGGANGTCTGACGTA | AGCTCTATGACCCGCTGT | ACACCCAGCATGTTCTGCGC | 553 | |
| QY | 459 | CGCGGAGGGCAGACCA | GAGACTCTGCAACGGTGA | CTGTGGGGGCCCTGATCTG | 518 | |
| DB | 554 | CGCGGAGGGCAGACCA | GAGACTCTGCAACGGTGA | CTGTGGGGGCCCTGATCTG | 613 | |
| QY | 519 | CAACGGGTACTGTCAGG | CCCTGTGTCTTTCGAAA | AGCCCGGTGTGCCAAT | TGGCCT 578 | |
| DB | 614 | CAACGGGTACTGTCAGG | CCCTGTGTCTTTCGAAA | AGCCCGGTGTGCCAAT | TGGCCT 673 | |
| QY | 579 | GCCAGGTGTCTACACCA | ACTCTGCAAAATTCATG | AGTGAGATAGAAAAAC | CGTCCAGNC 638 | |
| DB | 674 | GCCAGGTGTCTACACCA | ACTCTGCAAAATTCATG | AGTGAGATAGAAAAAC | CGTCCAGNC 733 | |
| QY | 639 | CAGTTACTCTGGGGACT | GGGAACCCATGAATTGAC | CCCCCAATACATCTGCG | GAANG 698 | |
| DB | 734 | CAGTTACTCTGGGGACT | GGGAACCCATGAATTGAC | CCCCCAATACATCTGCG | GAANG 793 | |
| QY | 699 | AATTAGGAATATCTGTT | CCAGCCCTCTCCTCAGG | CCCAAGAGTCCAGGCCCC | CAG 758 | |
| DB | 794 | AATTAGGAATATCTGTT | CCAGCCCTCTCCTCAGG | CCCAAGAGTCCAGGCCCC | CAG 853 | |
| QY | 759 | CCCCTCTCTCTCAAA | CAAGGATACAGATCCCA | AGCCCTCTCTCCTCAGAC | CCAGAG 818 | |
| DB | 854 | CCCCTCTCTCTCAAA | CAAGGATACAGATCCCA | AGCCCTCTCTCCTCAGAC | CCAGAG 913 | |
| QY | 819 | TCCAGACCCCAAGCCCT | CTCCTCAGACCCAGAG | AGTCCAGCCCTCTCTCCT | CAGAC 878 | |
| DB | 914 | TCCAGACCCCAAGCCCT | CTCCTCAGACCCAGAG | AGTCCAGCCCTCTCTCCT | CAGAC 973 | |
| QY | 879 | GCAGGAGTCCAGACCCC | CAGCCCTCCTGTCAGAC | CCCAAGGGGTGACGGCC | CCCAACC 938 | |

| | | | | | | |
|---|---|--------------------------|-------------------|-------------------|-------------------|----------|
| DB | 974 | CCAGAGTCCAGACCCCC | CAGCCCTCTCTCCTCA | GACCCAGGGGTGAGGCC | CCCAACC 1033 | |
| QY | 939 | CCTCCTCCTCAGAGTCA | GAGTCCAAAGCCCCCA | ACCCCTGCTTCCCAAG | ACAGAGT 998 | |
| DB | 1034 | CCTCCTCCTCAGAGTCA | GAGTCCAAAGCCCCCA | ACCCCTGCTTCCCAAG | ACAGAGT 1093 | |
| QY | 999 | NCAGTCCCAAGCCCTCT | CTCCCTCAGACCCAGC | GGGTCCATGACCCTAG | ANTTCCCTG 1058 | |
| DB | 1094 | NNAGTCCCAAGCCCTCT | CTCCCTCAGACCCAGC | GGGTCCATGACCCTAG | ANTTCCCTG 1153 | |
| QY | 1059 | TACACAGTGGCCCCCT | TGTGGCANGTTGACCC | CAACCTTACAGTTGTT | TGATTTTTG 1118 | |
| DB | 1154 | NACACAGTGGCCCCCT | TGTGGCANGTTGACCC | CAACCTTACAGTTGTT | TGATTTTTG 1213 | |
| QY | 1119 | TCCCTTCCCTAGATCCA | GAAATTAAGTNTAAGA | GAGCGCAAAAAA 1167 | | |
| DB | 1214 | TCCCTTCCCTAGATCCA | GAAATTAAGTNTAAGA | GAGCGCAAAAAA 1262 | | |
| RESULT 6 | | | | | | |
| ID | V61250 | standard; cDNA; 1265 BP. | | | | |
| AC | V61250; | | | | | |
| DT | 06-JAN-1999 | (first entry) | | | | |
| DE | cDNA sequence of prostate tumour clone P703 splice variant DE2. | | | | | |
| KW | Prostate; cancer; tumour; vaccine; immunogen; clone; ss. | | | | | |
| OS | Homo sapiens. | | | | | |
| PN | WO9837093-A2. | | | | | |
| PD | 27-AUG-1998. | | | | | |
| PF | 25-FEB-1998; U03492. | | | | | |
| PR | 09-FEB-1998; US-020956. | | | | | |
| PR | 25-FEB-1997; US-806099. | | | | | |
| PR | 01-AUG-1997; US-904804. | | | | | |
| PA | (CORI-) CORIXA CORP. | | | | | |
| PI | Dillon DC, Xu J; | | | | | |
| DR | WPI; 98-609886/51. | | | | | |
| PT | Polypeptides comprising immunogenic portions of prostate proteins - | | | | | |
| PS | used in a vaccine for the treatment of prostate cancer | | | | | |
| PS | Claim 3; Page 105-106; 130pp; English. | | | | | |
| CC | The present sequence is a new DNA which encodes an immunogenic portion | | | | | |
| CC | of a prostate tumour protein. The encoded immunogen, or the DNA itself, | | | | | |
| CC | can be used as a vaccine for the treatment of prostate cancer. The DNA | | | | | |
| CC | was identified by analysis of a subtracted cDNA library obtained by | | | | | |
| CC | subtracting a prostate tumour cDNA expression library with a normal | | | | | |
| CC | tissue cDNA library. | | | | | |
| SQ | Sequence | 1265 BP; | 256 A; | 432 C; | 321 G; | 245 T; |
| Query Match | | | | | | |
| Best Local Similarity 88.8%; Score 1036.2; DB 1; Length 1265; | | | | | | |
| Matches 1137; Conservative 0; Mismatches 29; Indels 83; Gaps 1; | | | | | | |
| QY | 2 | CGCAGCCCTGGCAGGGC | GCACGTGTCATGAAAAC | GAATTTGCTGCGGGCTC | TGG 61 | |
| DB | 14 | CGCAGCCCTGGCAGGGC | GCACGTGTCATGAAAAC | GAATTTGCTGCGGGCTC | TGG 73 | |
| QY | 62 | TGCATCCGACGTGGGTG | CTGTCAGCCGACACTGT | TCCAGAACTCCTACAC | CATCGGGC 121 | |
| DB | 74 | TGCATCCGACGTGGGTG | CTGTCAGCCGACACTGT | TCCAGAACTCCTACAC | CATCGGGC 133 | |
| QY | 122 | TGGGCTGCACAGTCTTG | AGCGCCGACCAAGACC | AGGAGCCAGATGGTGAG | GCCAGCC 181 | |
| DB | 134 | TGGGCTGCACAGTCTTG | AGCGCCGACCAAGACC | AGGAGCCAGATGGTGAG | GCCAGCC 193 | |
| QY | 182 | TCTCCGTACGGCACCC | AGATACACAGACTCTG | CTCGCTAACGACCTCA | TGCTCATCA 241 | |
| DB | 194 | TCTCCGTACGGCACCC | AGATACACAGACTCTG | CTCGCTAACGACCTCA | TGCTCATCA 253 | |
| QY | 242 | AGTTGACGAATCCGTGT | CCGAGTCTGACACCAT | CCGGAGCATCAGATTG | CTTCGCACT 301 | |
| DB | 254 | AGTTGACGAATCCGTGT | CCGAGTCTGACACCAT | CCGGAGCATCAGATTG | CTTCGCACT 313 | |
| QY | 302 | GCCCTACCGCGGGGA | ACTCTTGCCCTGTTCT | GGGTGTGCTGCGGAAC | ----- 355 | |
| DB | 314 | GCCCTACCGCGGGGA | ACTCTTGCCCTGTTCT | GGGTGTGCTGCGGAAC | GGGTGAGC 373 | |
| QY | 355 | ----- | ----- | ----- | 355 | |
| DB | 374 | TCACGGGTGTGTCTG | CCCTTCAAGAGAGTCT | CTGCCCCAATCGCGG | GGGTGACCC 433 | |
| QY | 355 | ----- | GGCAGATGCTTACCG | GTGCTGCACACTGCG | TGAACGTGCGTGT 398 | |
| DB | 434 | AGAGCTGTGCTCCAG | GCAAGATGCTTACCG | GTGCTGCACACTGCG | TGAACGTGCGTGT 493 | |
| QY | 399 | GTCAGGANGTCTGAC | GTAAGCTCTATGACCC | GCTGTACACCCAGCA | TGTTCTGCGC 458 | |
| DB | 494 | GTCAGGANGTCTGAC | GTAAGCTCTATGACCC | GCTGTACACCCAGCA | TGTTCTGCGC 553 | |
| QY | 459 | CGCGGAGGGCAGACA | GAGACTCTGCAACGG | TGACTGTGGGGGCC | CCCTGATCTG 518 | |
| DB | 554 | CGCGGAGGGCAGACA | GAGACTCTGCAACGG | TGACTGTGGGGGCC | CCCTGATCTG 613 | |
| QY | 519 | CAACGGGTACTGTC | AGGCTGTGTCTTTCG | AAAAGCCCGGTGTG | CCCAACTTGGCCT 578 | |
| DB | 614 | CAACGGGTACTGTC | AGGCTGTGTCTTTCG | AAAAGCCCGGTGTG | CCCAACTTGGCCT 673 | |
| QY | 579 | GCCAGGTGTCTACA | CCAACCTCTGCAAA | TTCACTGAGTGAGAT | AGAAAAACCGTCC | AGNC 638 |

Db 314 GCCCTACCGCGGGGAACCTTGGCTTCTGCTGGGCTCTGCTGGCAACGGTGAGC 373
QY 355 -----
Db 374 TCACGGGTGTGTCTGTCCCTCTTCAAGGAGGTCTCTGCCCAAGTCGCGGGGCTGACCC 433
QY 355 -----GGCAGAAATGCCCTACCGGTGCTGCACTGCTGAACGTGTGGTGTG 398
Db 434 AGAGCTGTGCGTCCAGGACAGAAATGCTTACCGGTGCTGCACTGCTGAACGTGTGGTGTG 493
QY 399 GTCTGAGGANGTCTGACGTATGACCCGCTGTACCAACCCAGCATGTCTGCGC 458
Db 494 GTCTGAGGANGTCTGACGTATGACCCGCTGTACCAACCCAGCATGTCTGCGC 553
QY 459 CGCGGAGGCGCAAGACCAAGAGACTCTGCAACGGTGTCTGCGGGGCGCCCTGATCTG 518
Db 554 CGCGGAGGCGCAAGACCAAGAGACTCTGCAACGGTGTCTGCGGGGCGCCCTGATCTG 613
QY 519 CAACGGGTACTTGACGGGCTGTGTCTTTCGGAAGGCGCGGTGTGGCAACTTGGCGT 578
Db 614 CAACGGGTACTTGACGGGCTGTGTCTTTCGGAAGGCGCGGTGTGGCAACTTGGCGT 673
QY 579 GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAACCGTCCAGNC 638
Db 674 GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAACCGTCCAGNC 733
QY 639 CAGTTAACTCTGGGACTGGGAACCCATGAATTTGACCCCAATACATCTCTGGGAANG 698
Db 734 CAGTTAACTCTGGGACTGGGAACCCATGAATTTGACCCCAATACATCTCTGGGAANG 793
QY 699 AATTGAGATATCTGTTCACAGCCCTCTCTCCAGGCGCCAGAGTCCAGGCGCCAG 758
Db 794 AATTGAGATATCTGTTCACAGCCCTCTCTCCAGGCGCCAGAGTCCAGGCGCCAG 853
QY 759 CCCCTCTCTCTCAAAACCAAGGGTACAGATCCCAAGCCCTCTCTCTCAAGACCAAGAG 818
Db 854 CCCCTCTCTCTCAAAACCAAGGGTACAGATCCCAAGCCCTCTCTCTCAAGACCAAGAG 913
QY 819 TCCAGACCCCGAGCCCT 878
Db 914 TCCAGACCCCGAGCCCT 973
QY 879 GCAGAGTCCAGACCCCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 938
Db 974 GCAGAGTCCAGACCCCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1033
QY 939 CCT 998
Db 1034 CCT 1093
QY 999 NCAGGTCCAGCCCT 1058
Db 1094 NCAGGTCCAGCCCT 1153
QY 1059 TACACAGTGGCCCT 1118
Db 1154 TACACAGTGGCCCT 1213
QY 1119 TCCCTTTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA 1167
Db 1214 TCCCTTTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA 1262

RESULT 7
V11855
ID V11855 standard; cDNA; 1386 BP.
AC V11855;
DT 11-SEP-1998 (first entry)
DE Homo sapiens Tub Interactor (hTI-1) gene.
KW serine protease; tub interactor; treatment; obesity; cachexia;
anorexia nervosa; diabetes; cell cycle progression; apoptosis;
neurodegenerative disease; Alzheimer's disease; drug screening;

KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 2..701
FT /product= hTI-1 protein
FT /note= "putative serine protease"
PN WO9812302-A1.
PD 26-MAR-1998.
PF 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errada PR; Gimeno CJ;
DR WPI; 98-217246/19.
DR P-PSDB; W59129.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 10; Fig 1; 120pp; English.
CC The sequence is that of the Tub Interactor gene hTI-1 which
CC codes for a putative serine protease. TI genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 81.4%; Score 950.2; DB 1; Length 1386;
Best Local Similarity 86.2%; Pred. No. 3.4e-228;
Matches 1132; Conservative 4; Mismatches 30; Indels 147; Gaps 3;

QY 2 CGACGCCCTGGCAGCGCGCACTGTGATGAAGAAAGCAATTTGCTGCGGCGCTCTCG 61
Db 21 CGACGCCCTGGCAGCGCGCACTGTGATGAAGAAAGCAATTTGCTGCGGCGCTCTCG 80
QY 62 TGCATCCGCACTGGGTGTGTCTGACGCCGACACACTGTTTCCAGAA-----CTCCT 109
Db 81 TGCATCCGCACTGGGTGTGTCTGACGCCGACACACTGTTTCCAGAGTGAAGTGCAGAGCTCT 140
QY 110 ACACCATCGGGCTGGGCTGTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 169
Db 141 ACACCATCGGGCTGGGCTGTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 200
QY 170 TGGAGGCCAGCCTCTCTCGTACGGCCACCCAGAGTACACAGACTCTTGCTCGCTAACGACC 229
Db 201 TGGAGGCCAGCCTCTCTCGTACGGCCACCCAGAGTACACAGACTCTTGCTCGCTAACGACC 260
QY 230 TCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCAATCCGAGCATCAGCA 289
Db 261 TCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTGACACCAATCCGAGCATCAGCA 320
QY 290 TTGCTTCGCACTGCCCTTACCGCGGGGAACCTTGTCTCTGCTGCGGCTGCTGCTGG 349
Db 321 TTGCTTCGCACTGCCCTTACCGCGGGGAACCTTGTCTCTGCTGCGGCTGCTGCTGG 380
QY 350 CGAAGCGCAGATGCTCTACCGTGTCTGCACTGCGTGAACGTGTGCTGTGAGGANG 409
Db 381 CGAAGCGCAGATGCTCTACCGTGTCTGCACTGCGTGAACGTGTGCTGTGAGGANG 440
QY 410 TCTGCACTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTCTGCGCGGCGGAGGGC 469
Db 441 TCTGCACTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTCTGCGCGGCGGAGGGC 500

| | | | |
|--------|--|---|------|
| QY | 470 | AAGACCAGAAGACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT | 529 |
| Db | 501 | AAGACCAGAAGACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT | 560 |
| QY | 530 | TGCAGGGCCTTGTGCTTTTCGAAAAAGCCCCCGTGTGGCCAATTGGCGTGCCAGGTGCT | 589 |
| Db | 561 | TGCAGGGCCTTGTGCTTTTCGAAAAAGCCCCCGTGTGGCCAATTGGCGTGCCAGGTGCT | 620 |
| QY | 590 | ACACCAACCTCTGCAAAATTCACGTGAGTGATAGAGAAAACCGT-CCAGNCCAGTTAATTC | 648 |
| Db | 621 | ACACCAACCTCTGCAAAATTCACGTGAGTGATAGAGAAAACCGTACCAGGCCAGTTAATTC | 680 |
| QY | 649 | TGGGGACTGGGAACCATGAAATTGACCCCCCAAATACATCTGCGGAANGAATTACAGAA | 708 |
| Db | 681 | TGGGGACTGGGAACCATGAAATTGACCCCCCAAATACATCTGCGGAANGAATTACAGAA | 740 |
| QY | 709 | TATCTGTTCCAGACCCCCTCTCCCTCAGGCCCAGAGATCCAGGCCCCCAGCCCTCTCTCC | 768 |
| Db | 741 | TATCTGTTCCAGACCCCCTCTCCCTCAGGCCCAGAGATCCAGGCCCCCAGCCCTCTCTCC | 800 |
| QY | 769 | CTCAAACCAAGGGGTACAGATCCCCAGCCCCCTCTCTCCCTCAGACCCAGAGATCCAGACCCC | 828 |
| Db | 801 | CTCAAACCAAGGGGTACAGATCCCCAGCCCCCTCTCTCCCTCAGACCCAGAGATCCAGACCCC | 860 |
| QY | 829 | CCAGCCCCCTCNTCCNTCAGACCCAGGAGTCCAGCCCCCTCTCCNTCAGACGACGAGATCC | 888 |
| Db | 861 | CCAGCCCCCTCTCTCCCTCAGACCCAGGAGTCCAGCCCCCTCTCTCCCTCAGACCCAGGATCC | 920 |
| QY | 889 | AGAC----- | 893 |
| Db | 921 | AGACCCCCCAGCCCCCTCTCTCTCAGACCCAGGGGTCCAGCCCTCTCTCTCAGACCCA | 980 |
| QY | 893 | ----- | 893 |
| Db | 981 | GGAGTCCAGACCCCCCAGCCCCCTCTCTCTCAGACCCAGGAGTCCAGCCCCCTCTCTCTC | 1040 |
| QY | 893 | -----CCCCCAGCCCNCTCCTGTCAGACCCAGGGGTGCAGGCCGCC | 934 |
| Db | 1041 | AGACCCAGGAGTCCAGATCCCCCAGCCCCCTCTCTCTCAGACCCAGGGGTCCAGGCCGCC | 1100 |
| QY | 935 | AACCCCTCNTCCNTCAGAGTCCAGAGTCCAGGCCCCCAACCCCTCGTTCCCCAGACCCAG | 994 |
| Db | 1101 | AACCCCTCTCTCTCAGACTCAGAGTCCAGGCCCCCAACCCCTCTCTCTCCCCAGACCCAG | 1160 |
| QY | 995 | AGGTNCAGGTCCACAGCCCCCTCTCTCTCTCAGACCCAGCCGCTCCAATGCCACCTAGANTTC | 1054 |
| Db | 1161 | AGGTCCAGGTACCAAGCCCCCTCTCTCTCTCAGACCCAGCCGCTCCAATGCCACCTATCTCTC | 1220 |
| QY | 1055 | CCTGTACACAGTGCCTCTGTGGCANGTTGACCCCAACCTTACCAGTTGTTTTCATTT | 1114 |
| Db | 1221 | CCTGTACANATTGCCNCTGTGGCACGTTGACCCCAACCTTACCAGTTGTTTTCATTT | 1280 |
| QY | 1115 | TTTGTCCTTTCCCTAGATCCAGAATAAAGTNTAAGAGAAGCGCAAAAAA | 1167 |
| Db | 1281 | TTTGTCCTTTCCCTAGATCCAGAATAAAGTNTAAGRGRAGSGCCAAAAA | 1333 |
| RESULT | 8 | | |
| ID | V37495 | standard; DNA; 871 BP. | |
| AC | V37495; | | |
| DT | 07-SEP-1998 | (first entry) | |
| DE | Human prostate-specific kallikrein (HPSK) encoding DNA. | | |
| KW | prostate-specific kallikrein; HPSK; prostate carcinoma; human; | | |
| OS | benign prostate hyperplasia; diagnosis; drug screening; PSK; ss. | | |
| FH | Homo sapiens. | | |
| Key | Location/Qualifiers | | |
| CDS | 31..777 | | |
| FT | /tag= a | | |
| FT | /transl_except= (pos:367..369, aa:Xaa) | | |
| FT | /transl_except= (pos:412..414, aa:Xaa) | | |
| FT | /transl_except= (pos:424..426, aa:Xaa) | | |
| FT | /product= "HPSK protein" | | |

| | | | |
|--|---|------------------------|--|
| FT | | /note= "Xaa - unknown" | |
| PN | W09820117-A1. | | |
| PD | 14-MAY-1998. | | |
| PF | 31-OCT-1997; U20051. | | |
| PR | 05-NOV-1996; US-744026. | | |
| PA | (INCY-) INCYTE PHARM INC. | | |
| PI | Bandman O, Gol S; | | |
| DR | WPI; 98-286933/25. | | |
| P-PSDB: | W60592. | | |
| PT | New isolated prostate-specific kallikrein - used to develop products | | |
| PT | for diagnosis and treatment of, e.g. prostate carcinoma or benign | | |
| PT | hyperplasia | | |
| PS | Claim 5; Fig 1A-C; 68bp; English. | | |
| CC | This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell | | |
| CC | containing an expression vector comprising the HPSK nucleic acid sequence | | |
| CC | can be used to produce the protein recombinantly. The HPSK products can | | |
| CC | be used for the diagnosis of conditions or diseases associated with | | |
| CC | expression of HPSK such as prostate carcinoma and benign prostate | | |
| CC | hyperplasia. Agonists and antagonists which specifically bind to HPSK and | | |
| CC | modulate its activity can be used for the preparation of treatment of | | |
| CC | such conditions or diseases. The products can also be used for detection | | |
| CC | and drug screening, especially for the detection of prostate-specific | | |
| CC | kallikrein (PSK). | | |
| SQ | Sequence 871 BP; 166 A; 260 C; 258 G; 184 T; | | |
| | | | |
| Query Match 61.8%; Score 721.4; DB 1; Length 871; | | | |
| Best Local Similarity 98.2%; Pred. No. 3.6e-171; | | | |
| Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0 | | | |
| | | | |
| QY | 2 CGCAGCCCTGGCAGGCGGCACACTGGTCATGGAAAACGAATTGTCTGCTCGGCCCTCTGG | 61 | |
| Db | 134 CCCAGCCCTGGCAGGCGGCACACTGGTCATGGAAAACGAATTGTCTGCTCGGCCCTCTGG | 193 | |
| | | | |
| QY | 62 TGCAATCCGCACTGGTGTCTGTACGCGCACACTGTTTCCAGAATCTTACACCATCGGGC | 121 | |
| Db | 194 TGCATCCGCACTGGTGTCTGTACGCGCACACTGTTTCCAGAATCTTACACCATCGGGC | 253 | |
| | | | |
| QY | 122 TGGSCCTGCACAGTCTTGAGGCCGACAAGACCAGGAGCCAGATGGTGGAGCCAGCC | 181 | |
| Db | 254 TGGSCCTGCACAGTCTTGAGGCCGACAAGACCAGGAGCCAGATGGTGGAGCCAGCC | 313 | |
| | | | |
| QY | 182 TCTCCGTACGGCACCAGAGTACAAACAGACTTGTCTCGCTAACGACCTCATGCTCATCA | 241 | |
| Db | 314 TCTCCGTACGGCACCAGAGTACAAACAGACTTGTCTCGCTAACGACCTCATGNTCATCA | 373 | |
| | | | |
| QY | 242 AGTTGGACGAATCCGTGTCCGAGTGTACACACCATCCGGAGCATGATGCTTGGCAGT | 301 | |
| Db | 374 AGTTGGACGAATCCGTGTCCGAGTGTACACACATCCGGAGNATGATGNTTGGCAGT | 433 | |
| | | | |
| QY | 302 GCCCTACCGGGGGGAACCTTGGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 361 | |
| Db | 434 GCCCTACCGGGGGGAACCTTGGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 493 | |
| | | | |
| QY | 362 TGCCTACCGGTGTGTGCACTGCGTGAACGTTGCGGTGTGTGAGGANGTCTGCAGTAAGC | 421 | |
| Db | 494 TGCCTACCGGTGTGTGCACTGCGTGAACGTTGCGGTGTGTGAGGANGTCTGCAGTAAGC | 553 | |
| | | | |
| QY | 422 TCATATGACCCGCTGTATCCACCCCAGCATGTTCTGCGCCGCGCGGCAAGAGCAGAAAG | 481 | |
| Db | 554 TCATATGACCCGCTGTATCCACCCCAGCATGTTCTGCGCCGCGGCGGCAAGAGCAGAAAG | 613 | |
| | | | |
| QY | 482 ACTCCTGCAACGCTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGACAGGCTTG | 541 | |
| Db | 614 ACTCCTGCAACGCTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGACAGGCTTG | 673 | |
| | | | |
| QY | 542 TGTCTTTGGAAAAAGCCCCGTGTGGCCAACCTTGGCGTGCCAGGTGTCTACACCAACTCT | 601 | |
| Db | 674 TGTCTTTGGAAAAAGCCCCGTGTGGCCAAGTTGGCGTGCCAGGTGTCTACACCAACTCT | 733 | |
| | | | |
| QY | 602 GCAAATTCAGTAGTGATAGAGAAAACCGTCCAGNCCAGTTAACTCTGGGAGCTGGGAA | 661 | |
| Db | 734 GCAAATTCAGTAGTGATAGAGAAAACCGTCCAGGAGCCAGTTAACTCTGGGAGCTGGGAA | 793 | |

| | | | |
|----|-----|---|-----|
| OY | 662 | CCCCATGAATTGACCCCAATAATACATCCTCGGGGAANGAATTCAGAAATATCTGTGCCAG | 721 |
| | | | |
| | | | |
| | | | |
| | | | |
| Dd | 794 | CCCATGAATTGACCCCAATAATACATCCTCGGGGAAGAATTCAGAAATATCTGTGCCAG | 853 |
| | | | |
| | | | |
| | | | |
| | | | |
| Oy | 722 | CCCCCTCCTCCCTCAGGCC | 739 |
| | | | |
| | | | |
| | | | |
| Dd | 854 | CCCCCTCCTCCCTCAGGCC | 871 |

RESULT 9

| | |
|----------|---|
| ID | V58648 standard; CDNA; 1119 BP. |
| AC | V58648; |
| DT | 08-DEC-1998 (first entry) |
| DE | Prostate tumour specific gene clone DE14. |
| KW | Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss. |
| OS | Homo sapiens. |
| FH | Key |
| FT | CDS |
| PN | WO9837418-A2. |
| PD | 27-AUG-1998. |
| PF | 25-FEB-1998; U03690. |
| PR | 09-FEB-1998; US-904809. |
| PR | 25-FEB-1997; US-806596. |
| PR | 01-AUG-1997; US-904809. |
| PA | (CORI-) CORIXA CORP. |
| PI | Dillon DC, Xu J; |
| DR | WPI; 98-480805/41. |
| DR | P-PSTDB; W69389. |
| PT | Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers |
| PS | Claim 1: Page 116-117; 141pp; English. |
| CC | This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers. |
| CC | |
| CC | |
| CC | |
| Sequence | 1119 BP; 248 A; 305 C; 282 G; 284 T; |

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 30.2%; | Score 352.2; | DB 1; | Length 1119; |
| Best Local Similarity | 96.3%; | Pred. No. 4.8e-79; | | |
| Matches 360; Conservative | 0; | Mismatches 14; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| OY | 2 | CGCAGCCCTGGCAGGCGGCACCTGGTCATGGAACAAGAAATTGTTCGCTCGGCCGTCTCG | 61 |
| Db | 8 | CGCAGCCCTGGCAGGCGGCACCTGGTCATGGAACAAGAAATTGTTCGCTCGGCCGTCTCG | 67 |
| OY | 62 | TGCATCCGCACTGGGTGCTGTCAAGCCGACACTGTTCCAGAATCCTACACCATCGGC | 121 |
| Db | 68 | TGCATCCGCACTGGGTGCTGTCAAGCCGACACTGTTCCAGAATCCTACACCATCGGC | 127 |
| OY | 122 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGAGGCCAGCC | 181 |
| Db | 128 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGAGGCCAGCC | 187 |
| OY | 182 | TCTCCGTACGGCACCCAGAGTACACACAGACTTGTCTCGTAACGACCTCATGTCTATCA | 241 |
| Db | 188 | TCTCCGTACGGCACCCAGAGTACACACAGACCTTGTCTCGTAACGACCTCATGTCTATCA | 247 |
| OY | 242 | AGTTGGACGAATCCGTGTCGGAGTGTGACACCATCCGGAGCATCAGCATTTGCTTCGAGT | 301 |
| Db | 248 | AGTTGGACGAATCCGTGTCGGAGTGTGACACCATCCGGAGCATCAGCATTTGCTTCGAGT | 307 |
| OY | 302 | GCCCTACCGCGGGGAATCTTGCCTCGTNTTCGGCTGGGCTGTGCTGGCGAAGCGCAGAA | 361 |

Db 308 GCCCTACCGCGGGGGAACCTCTTGCCTCGTTCTGGCTGGGGTCTGCTGGCGAAGATGCTG 367

QY 362 TGCCCTACCGTGCCTG 375

Db 368 TGATTGCCATCCAG 381

RESULT 10

ID V61253 standard; cDNA: 1119 BP.
 AC V61253;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN W09837093-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; U03492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 PS P-PSDB; W71873.
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 PS Claim 3; Page 108-109; 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 CC Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;
 SQ

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 30.2%; | Score 352.2; | DB 1; | Length 1119; |
| Best Local Similarity | 96.3%; | Pred. No. 4.8e-79; | | |
| Matches 360; | Conservative 0; | Mismatches 14; | Indels 0; | Gaps 0; |

[illegible]

RESULT 11

X41114
ID X41114 standard; cDNA; 402 BP.

AC X41114; (first entry)
DT 17-JUN-1999 EST SEQ ID NO:58.
DE Human secreted protein 5' EST
KW Human: secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemokine; chemokine; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR P-PSDB; Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide or the insertion of a polypeptide into a membrane, or
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 28.4%; Score 331.4; DB 1; Length 402;
Best Local Similarity 96.2%; Pred. No. 5.5e-74;
Matches 354; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

OY 2 CGCAGCCCTGGCAGCGGCACTGTCATGAAACGAATGTTCTGCTGGGCTCCTGG 61
DB 29 CGCAGCCCTGGCAGCGGCACTGTCATGAAACGAATGTTCTGCTGGGCTCCTGG 88
OY 62 TGCATCCGAGTGGTCTGTCAGCCGACACTGTTCCAGAA-----CTCCT 109
DB 89 TGCATCCGAGTGGTCTGTCAGCCGACACTGTTCCAGAAAGTGTGCAGAGCTCCT 148
OY 110 ACACCATCGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCGGAGCCAGATGG 169
DB 149 ACACCATCGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCGGAGCCAGATGG 208
OY 170 TGGAGGCCAGCCTCCGTACGCGACCCAGAGTACACAGAGCTTGTCTCGCTAACGACC 229
DB 209 TGGAGGCCAGCCTCCGTACGCGACCCAGAGTACACAGAGCTTGTCTCGCTAACGACC 268
OY 230 TCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGGAGCATCAGCA 289
DB 269 TCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGGAGCATCAGCA 328
OY 290 TTGCTTCGAGTCCCTACCGCGGGGAATCTTGCTGCTGCTGGGCTGCTGCTGG 349
DB 329 TTGCTTCGAGTCCCTACCGCGGGGAATCTTGCTGCTGCTGGGCTGCTGCTGG 388
OY 350 CGAAGCGC 357
DB 389 CGAAGCGC 396

RESULT 12
V58646 ID V58646 standard; cDNA; 1459 BP.
AC V58646;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 25.1%; Score 292.8; DB 1; Length 1459;
Best Local Similarity 80.4%; Pred. No. 3.4e-64;
Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

OY 92 ACTGTTCCAGAACTCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCGCAAG 151
DB 25 AGTAGTGACAGAGCTCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCGCAAG 84
OY 152 AGCCAGGAGCAGATGTGAGGCCAGCCTCTCCGTACGCCACCCAGAGTACACAGAC 211
DB 85 AGCCAGGAGCAGATGTGAGGCCAGCCTCTCCGTACGCCACCCAGAGTACACAGAC 144
OY 212 TCTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCCAGTGTGACA 271
DB 145 CTTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCCAGTGTGACA 204
OY 272 CCATCCGGAGCATCAGATTGCTTCGAGTGCCCTACCGGGGAACTCTTGCTGCTNT 331
DB 205 CCATCCGGAGCATCAGATTGCTTCGAGTGCCCTACCGGGGAACTCTTGCTGCTNT 264
OY 332 CTGGCTGGGCTGCTGGGCAAC-----GGCAGAAATGCTTAC 355
DB 265 CTGGCTGGGCTGCTGGGCAACGCTGAGCTCAGGGGTGTGTCTGCCCTCTTCAAGGA 324
OY 355 -----GGCAGAAATGCTTAC 368
DB 325 GGTCTCTGCCAGTCCGGGGCTGACCCAGAGCTCTGCTGCCAGCAGAAATGCTTAC 384
OY 369 CGTGTGACATCGCTGACAGCTGTGCTGCTGCTGAGGANGTCTGCAGTAAGCTATGA 428
DB 385 CGTGTGACATCGCTGACAGCTGTGCTGCTGCTGAGGANGTCTGCAGTAAGCTATGA 444
OY 429 CCCGCTGATCAACCCAGCATGTTCTGCGCGGGGAGGCAAGACCAAGAGACTCTG 488
DB 445 CCCGCTGATCAACCCAGCATGTTCTGCGCGGGGAGGCAAGACCAAGAGACTCTG 504
OY 489 CAACG 493
DB 489 CAACG 493

Db 505 CAACG 509

RESULT 13

ID V61251 standard; cDNA; 1459 BP.

AC V61251;

DT 06-JAN-1999 (first entry)

DE cDNA sequence of prostate tumour clone P703 splice variant DE6.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens.

PN WO9837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins -

PS Claim 3; Page 106; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA was identified by analysis of a subtracted cDNA library obtained by CC subtracting a prostate tumour cDNA expression library with a normal CC tissue cDNA library.

SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 25.1%; Score 292.8; DB 1; Length 1459;

Best Local Similarity 80.4%; Pred. No. 3.4e-64;

Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 92 ACTGTTCCAGAACTCCTACACCATCGGGCTGGCTGCACAGTCTTGAGCCGACCAAG 151

Db 25 AGTGAGTGACAGAGCTCCTACACCATCGGGCTGGCTGCACAGTCTTGAGCCGACCAAG 84

QY 152 AGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 211

Db 85 AGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGAC 144

QY 212 TCTGTGCTGCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTGTCCAGTCTGACA 271

Db 145 CCTGTGCTGCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTGTCCAGTCTGACA 204

QY 272 CCATCCGGAGCATCAGCATGCTTCCAGTGGCCCTACCGCGGGAACCTTGCCTCGTNT 331

Db 205 CCATCCGGAGCATCAGCATGCTTCCAGTGGCCCTACCGCGGGAACCTTGCCTCGTNT 264

QY 332 CTGGCTGGGGTCTGTGGCGAAC----- 355

Db 265 CTGGCTGGGGTCTGTGGCGAACGGTGTGTGTCTGCCCTCTTCAAGGA 324

QY 355 -----GGCAGATGCTTAC 368

Db 325 GGTCCTCTGCCAGTCCGGGGGCTGACCCAGAGCTCTGCCCTCCAGGCGAGATGCTTAC 384

QY 369 CGTGCTGCACTGCGTGAACGTGTGGGTGTGTGTGAGGANGTCTGCAGTAACTCTATGA 428

Db 385 CGTGCTGCACTGCGTGAACGTGTGGGTGTGTGTGAGGANGTCTGCAGTAACTCTATGA 444

QY 429 CCCGCTGTACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCGCAAGACCAGAGGACTCTCTG 488

Db 445 CCCGCTGTACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCGCAAGACCAGAGGACTCTCTG 504

QY 489 CAACG 493

Db 505 CAACG 509

RESULT 14

ID V58522 standard; cDNA; 234 BP.

AC V58522;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone P20.

KW Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss.

OS Homo sapiens.

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-480805/41.

PT Novel human prostate specific tumour protein and fragments - useful

PS Claim 1; Page 56; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting CC prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as CC encoded by this sequence). An antibody which binds to an immunogenic CC portion of the prostate protein, and the method can be used to detect, CC monitor progression of, or treat prostate cancers. The antibody may CC also be conjugated to a therapeutic agent for use in therapy of prostate CC cancers.

SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 19.6%; Score 228.8; DB 1; Length 234;

Best Local Similarity 98.3%; Pred. No. 1.9e-48;

Matches 230; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 203 ACAACAGACTCTTGTGCTGCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTCCG 262

Db 1 ACAACAGACCTTGTGCTGCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTCCG 60

QY 263 AGTCTGACACCATCCGAGAGCATCAGCATGCTTGCAGTGGCCCTACCGCGGGAACCTTT 322

Db 61 AGTCTGACACCATCCGAGAGCATCAGCATGCTTGCAGTGGCCCTACCGCGGGAACCTTT 120

QY 323 GCCTGCTNTCTGGCTGGGGTCTGCTGGCGCAACGGCAGAAATGCTTACCGTGTGCTACCTCGG 382

Db 121 GCCTGCTTCTGGCTGGGGTCTGCTGGCGCAACGGCAGAAATGCTTACCGTGTGCTACCTCGG 180

QY 383 TGAACGTGTGGGTGTGTGAGGANGTCTGAGTAAGCTCTATGACCCGCTGT 436

Db 181 TGAACGTGTGGGTGTGTGAGGANGTCTGAGTAAGCTCTATGACCCGCTGT 234

RESULT 15

ID V61168 standard; cDNA; 234 BP.

AC V61168;

DT 06-JAN-1999 (first entry)

DE cDNA sequence of prostate tumour clone P20.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens.

PN WO9837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins -

PS Claim 3; Page 106; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA was identified by analysis of a subtracted cDNA library obtained by CC subtracting a prostate tumour cDNA expression library with a normal CC tissue cDNA library.

PS Claim 3; Page 53-54; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 19.6%; Score 228.8; DB 1; Length 234;
Best Local Similarity 98.3%; Pred. No. 1.9e-48;
Matches 230; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 203 ACAACAGACTCTGCTCGCTACGACCTCATGCTCATCAAGTTGAGCGAATCCGTGTCG 262
Db 1 ACAACAGACCCCTTGCTGCTAGCAGACCTCATGCTCATCAAGTTGAGCGAATCCGTGTCG 60
QY 263 AGTCTGACACACATCCGGAGCATCAGCATGCTTGCAGTGCCTTACCCGGGGAACCTCT 322
Db 61 AGTCTGACACACATCCGGAGCATCAGCATGCTTGCAGTGCCTTACCCGGGGAACCTCT 120
QY 323 GCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGGAGATGCCCTACCGTGTGCACCTGC 382
Db 121 GCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGGAGATGCCCTACCGTGTGCACCTGC 180
QY 383 TGAACGTGTGGTGTCTGAGGANGTCTGCAAGTAAGCTCTATGACCCGCTGT 436
Db 181 TGAACGTGTGGTGTCTGAGGANGTCTGCAAGTAAGCTCTATGACCCGCTGT 234

Search completed: September 28, 1999, 12:28:42
Job time: 5342 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:33:47 ; Search time 161.06 Seconds
(without alignments)
664.265 Million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167
Sequence: 1 GCGCAGCCCTGGCAGCGCGC.....NTAAGAGAAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-------------------|--------------------|
| 1 | 721.4 | 61.8 | 871 | 3 | US-08-744-026-2 | Sequence 2, Appl1 |
| 2 | 181.6 | 15.6 | 8174 | 1 | US-07-914-281-5 | Sequence 5, Appl1 |
| 3 | 181.6 | 15.6 | 8174 | 1 | US-08-393-246-5 | Sequence 5, Appl1 |
| 4 | 181.6 | 15.6 | 8174 | 3 | US-08-525-058A-5 | Sequence 5, Appl1 |
| 5 | 181.6 | 15.6 | 8174 | 5 | PCT-US91-00899-3 | Sequence 3, Appl1 |
| 6 | 154.6 | 13.2 | 1089 | 5 | PCT-US96-04294-1 | Sequence 1, Appl1 |
| 7 | 154.6 | 13.2 | 1089 | 5 | PCT-US96-04294-3 | Sequence 3, Appl1 |
| 8 | 153 | 13.1 | 986 | 4 | US-08-557-146-1 | Sequence 1, Appl1 |
| 9 | 142.2 | 12.2 | 4060 | 1 | US-08-308-949A-1 | Sequence 1, Appl1 |
| 10 | 140 | 12.0 | 732 | 3 | US-08-361-395-2 | Sequence 2, Appl1 |
| 11 | 125.4 | 10.7 | 833 | 4 | US-08-790-137-2 | Sequence 2, Appl1 |
| 12 | 95.8 | 8.2 | 957 | 3 | US-08-684-862-11 | Sequence 11, Appl1 |
| 13 | 85.8 | 7.4 | 1462 | 2 | US-08-358-782D-14 | Sequence 14, Appl1 |
| 14 | 84.2 | 7.2 | 992 | 2 | US-08-358-782D-13 | Sequence 13, Appl1 |
| 15 | 84.2 | 7.2 | 1729 | 4 | US-08-844-024-1 | Sequence 1, Appl1 |
| 16 | 81 | 6.9 | 832 | 5 | PCT-US95-06157-5 | Sequence 5, Appl1 |
| 17 | 81 | 6.9 | 760 | 5 | PCT-US95-06157-7 | Sequence 7, Appl1 |
| 18 | 81 | 6.9 | 766 | 5 | PCT-US95-06157-9 | Sequence 9, Appl1 |
| 19 | 80 | 6.9 | 1454 | 3 | US-08-467-155A-2 | Sequence 2, Appl1 |
| 20 | 80 | 6.9 | 1454 | 4 | US-08-628-198-2 | Sequence 2, Appl1 |
| 21 | 80 | 6.9 | 1454 | 5 | PCT-US96-07343-2 | Sequence 8, Appl1 |
| 22 | 75.4 | 6.5 | 699 | 3 | US-08-738-413B-8 | Sequence 14, Appl1 |
| 23 | 71.2 | 6.1 | 840 | 3 | US-08-684-862-12 | Sequence 12, Appl1 |
| 24 | 70.4 | 6.0 | 1333 | 3 | US-08-684-862-9 | Sequence 9, Appl1 |
| 25 | 66 | 5.7 | 988 | 3 | US-08-684-862-10 | Sequence 10, Appl1 |
| 26 | 65.6 | 5.6 | 1096 | 3 | US-08-684-862-8 | Sequence 8, Appl1 |
| 27 | 62.4 | 5.3 | 7218 | 2 | US-08-232-463-14 | Sequence 14, Appl1 |
| 28 | 57.8 | 5.0 | 734 | 3 | US-08-650-129-1 | Sequence 1, Appl1 |
| 29 | 57.8 | 5.0 | 821 | 3 | US-08-650-129-2 | Sequence 2, Appl1 |
| 30 | 57.8 | 5.0 | 866 | 3 | US-08-650-129-3 | Sequence 3, Appl1 |
| 31 | 54 | 4.6 | 2259 | 4 | US-08-845-998-3 | Sequence 3, Appl1 |
| 32 | 52.4 | 4.5 | 2218 | 4 | US-08-845-998-5 | Sequence 5, Appl1 |
| 33 | 51.4 | 4.4 | 200 | 3 | US-08-931-981A-1 | Sequence 1, Appl1 |
| 34 | 49 | 4.2 | 1554 | 3 | US-08-469-486-1 | Sequence 1, Appl1 |
| 35 | 47.2 | 4.0 | 8174 | 1 | US-07-914-281-5 | Sequence 5, Appl1 |
| 36 | 47.2 | 4.0 | 8174 | 1 | US-08-393-246-5 | Sequence 5, Appl1 |
| 37 | 47.2 | 4.0 | 8174 | 3 | US-08-525-058A-5 | Sequence 5, Appl1 |

| | | | | | | |
|----|------|-----|------|---|------------------|--------------------|
| 38 | 47.2 | 4.0 | 8174 | 5 | PCT-US91-00899-3 | Sequence 3, Appl1 |
| 39 | 46.8 | 4.0 | 884 | 4 | US-08-851-974-2 | Sequence 2, Appl1 |
| 40 | 46 | 3.9 | 329 | 1 | US-08-148-910-2 | Sequence 2, Appl1 |
| 41 | 46 | 3.9 | 970 | 1 | US-08-148-910-3 | Sequence 3, Appl1 |
| 42 | 46 | 3.9 | 329 | 1 | US-08-148-910-13 | Sequence 13, Appl1 |
| 43 | 46 | 3.9 | 2033 | 1 | US-08-148-910-14 | Sequence 14, Appl1 |
| 44 | 46 | 3.9 | 329 | 2 | US-08-448-937A-2 | Sequence 2, Appl1 |
| 45 | 46 | 3.9 | 970 | 2 | US-08-448-937A-3 | Sequence 3, Appl1 |

ALIGNMENTS

RESULT 1
US-08-744-026-2
; Sequence 2, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-744-026-2

Query Match 61.8%; Score 721.4; DB 3; Length 871;
Best Local Similarity 98.2%; Pred. No. 1.2e-184;
Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 CCGAGCCCTGGCAGCGCGCATGTCATGGAACGAATGTCTGCTCGGCGTCTGG 61
Db 134 CCGAGCCCTGGCAGCGCGCATGTCATGGAACGAATGTCTGCTCGGCGTCTGG 193
QY 62 TGCATCGCAGTGGGTGCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATGGGGC 121
Db 194 TGCATCGCAGTGGGTGCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATGGGGC 253


```
QY 122 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGTGAGGCCAGCC 181
Db 254 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGTGAGGCCAGCC 313
QY 182 TCTCCGTACGGCAGCCAGAGTACAAACAGACTCTGCTCGCTAACGACCTCATGCTCATCA 241
Db 314 TCTCCGTACGGCAGCCAGAGTACAAACAGACTCTGCTCGCTAACGACCTCATGCTCATCA 373
QY 242 AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATCTGCTCCAGT 301
Db 374 AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATCTGCTCCAGT 433
QY 302 GCCCTACCGCGGGGAGACTCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGGAGAA 361
Db 434 GCCCTACCGCGGGGAGACTCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGGAGAA 493
QY 362 TGCCTACCGCTGCTGACACTGCGGTGAACGCTGCGGTGCTGCTGAGGANGTCTGCACTAAGC 421
Db 494 TGCCTACCGCTGCTGACACTGCGGTGAACGCTGCGGTGCTGCTGAGGANGTCTGCACTAAGC 553
QY 422 TCTATGACCCGCTGTACCAACCCAGCATGTCTGCGCGCGGAGGAGGAGCAGAGAGG 481
Db 554 TCTATGACCCGCTGTACCAACCCAGCATGTCTGCGCGCGGAGGAGGAGCAGAGAGG 613
QY 482 ACTCCTGCAACGCTGACTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 541
Db 614 ACTCCTGCAACGCTGACTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 673
QY 542 TGTCTTTTCGGAAGAGCCCGGTGTGGCCAACTTGCGCTGCGAGGTGTCTACCAACCACTCT 601
Db 674 TGTCTTTTCGGAAGAGCCCGGTGTGGCCAACTTGCGCTGCGAGGTGTCTACCAACCACTCT 733
QY 602 GCAAAATTCAGTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
Db 734 GCAAAATTCAGTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 793
QY 662 CCCATGAATTTGACCCCAATATCATCTGCGGGAAGAATTCAGGAATATCTGTTCCAG 721
Db 794 CCCATGAATTTGACCCCAATATCATCTGCGGGAAGAATTCAGGAATATCTGTTCCAG 853
QY 722 CCCCTCTCTCCCTCAGGCC 739
Db 854 CCCCTCTCTCCCTCAGGCC 871
```

RESULT 2
US-07-914-281-5/c
Sequence 5, Application US/07914281
Patent No. 5324663

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-914-281-5

Query Match 15.6%; Score 181.6; DB 1; Length 8174;
Best Local Similarity 77.3%; Pred. No. 1.2e-39;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

```
QY 700 ATTCAGATATATCTGTCCAGAGCCGCTCTCTCAGGAGCCAGAGTCCAGGAGCCAGC 759
Db 1025 ACTTAGAGTTCAGAGCTCCGCGCCCTCTCTCTCAGAGCCAGAGTCCAGAGCCAGC 966
QY 760 CCTCTCTCTCTCAACCAAGG-GTACAGATCCCAAGCCCTCTCTCTCTCAGAGCCAGAG 818
Db 965 CCTCTCTCTCTCAACCAAGAGTCCAGAGCCCAAGCCCTCTCTCTCTCAGAGCCAGAG 906
QY 819 TCCAGAGCCCGGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 870
Db 905 TCCAGAG-TCCCTAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847
QY 871 CNTCAGAGCAGAGAGTCCAGAGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 930
Db 846 CCTCAGAGCCAGAGAGCCCAAGTCTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787
QY 931 CCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 990
Db 787 -CCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729
QY 991 CCAGAGTNCAGGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1038
Db 728 CCAGAGTCCAGAGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 681
```

RESULT 3
US-08-393-246-5/c
Sequence 5, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye Ph.D., Jean-Paul
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4686..5780
OTHER INFORMATION: /label= mat_peptide
PCT-US91-00899-3

Query Match 15.6%; Score 181.6; DB 5; Length 8174;
Best Local Similarity 77.3%; Pred. No. 1.2e-39;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

OY 700 ATTGAGGATATCTGTTCCTCCAGCCCTCCTCCTCAGGCCAGGAGTCCAGGCCCCAGC 759
DB 1025 ACTTAGAGTCCAGGCTCCGCGCCCTCCTCCTCAGACCAGAGTCCAGGCCCTGC 966
OY 760 CCCTCCTCCTCAACAAGG-GTACAGATCCCGCCCTCCTCCTCAGACCAGAG 818
DB 965 CCCTCCTCCTCAGACCAGAGTCCAGACCAGGCCCTCCTCCTCAGACCAGAG 906
OY 819 TCCAGACCCCCAGCCCTCCTCCTCAGACCAGAGTCCAG-----CCCTCCTC 870
DB 905 TCCAGA-TCCCTAGCCCTACTCCTCAGACCAGAGTCCAGACCAGTCCCTCCTC 847
OY 871 CNTCAGACCCAGAGTCCAGACCCCCAGCCCTCCTCAGACCAGAGGAGTCCAGGC 930
DB 846 CCTCAGACCCAGAGGCCAGAGTCCAGGCCCTCCTCCTCAGATCCAGAGTACAGG- 787
OY 931 CCCCAACCCCTCCTCCTCAGAGTCCAGAGTCCAGGCCCAACCCCTCCTCCTCAGAC 990
DB 787 -CCAGACCCCTCCTCCTCAGAGTCCAGAGTCCAGGCCCAACCCCTCCTCCTCAGAC 729
OY 991 CCAGAGTTCAGAGTCCAGGCCCTCCTCCTCAGACCAGAGGAGTCCAA 1038
DB 728 CCAGAGTTCAGAGGCCAGGCCCTCCTCCTCAGACACAGAGGAGGCTA 681

RESULT 6
PCT-US96-04294-1/c
Sequence 1, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.

TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bialock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1089 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US96-04294-1

Query Match 13.2%; Score 154.6; DB 5; Length 1089;
Best Local Similarity 55.1%; Pred. No. 9.7e-33;
Matches 351; Conservative 0; Mismatches 271; Indels 15; Gaps 2;

OY 2 CGCAGCCCTGGCAGCGGCACTGTCATGGAAGAATGTTCTGCGGCGTCTG 61
DB 855 CCCACCATGCGAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGGGAGGCGTCTG 796
OY 62 TGCATCCGCAAGTGGTGTCTCAGCCGACACTGTTCCAGAACTCTACACCATCGGGC 121
DB 795 TCATGAGCGCTGGGTGCTCAGTCCGCCCTCAGATGAATGATGATACACCGTGCACC 736
OY 122 TGGGCTGCACAGTCTTGAAGCCGACCAAGAGCAGGAGCCAGATGTTGAGGCCAGCC 181
DB 735 TGGGCAGTGATACGCTGGGCGACAGAGAG-----CTCAGAGGATCAAGGCCCTCGA 685
OY 182 TCTCGTACGGCAGCCAGAGTACACAGACTCTGCTCGCTAACGAGCTCATGCTCATCA 241
DB 684 AGTATTTCCGCGCAGCCGCTACTCCACACAGACCATGTTAATGAGCTCATGCTGTA 625
OY 242 AGTTGACGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATGCTTGCAGT 301
DB 624 AGCTCAATAGCCAGGCGCAGCTGTCATCCATGTTGAAGAAGTCAAGGCTGCCCTCCGCT 565
OY 302 GCCCTACCGCGGGAAGTCTGCTCGTCTGCTGCGGCTGCTGCTGCGGAAC-----G 355
DB 564 GCGAACCCCTGGAACCACTGTACTGTCTCCGCTGGGGCACTACACGAGCCAGATG 505
OY 356 GCAATATGCCCTACCGTGTCTGCACTGCGTGAACGTGTGGTGTCTGAGGANGTCTGCA 415
DB 504 TGACCTTTCCCTCTGACTCATGTGCGTGAATGTCAGCTCATCTCCGCCAGGACTGCA 445
OY 416 GTAAGCTTATGACCGCGTGTACCAACCCAGCATGTTCTGCGCGCGGAGGAGGCAAGACC 475

Db 444 CGAAGTTTACAAGACTTACTGGAATAATTCATGCTGTGCGGTGGCATCCCGACTCCA 385
QY 476 AGAAGACTCCTGCAACGGTACTCTGGGGCCCTGATCTGCAACGGGTACTTGACAG 535
Db 384 AGAAAAAGCCTGCAATGCTGACTCAGGGGGACCGTTGTTGTCAGAGGTACCTGCAAG 325
QY 536 GCCTTGTGTCTTTCGAAAAAGCCCCGTGTGGCCAACCTGGCGTGGCAGGTGTCTACACCA 595
Db 324 GTCTGGTGTCTTGGGGAACCTTCCCTTGGCGGCAACCAATGACCCAGAGTCTACACTC 265
QY 596 ACCTCTGCAATTCACCTGAGTGGATAGAGAAAACCGT 632
Db 264 AAGTCTGCAAGTTCACCAAGTGATTAATGACACCAT 228

RESULT 7

PCT-US96-04294-3/c
; Sequence 3, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; PCT-US96-04294-3

Query Match 13.2%; Score 154.6; DB 5; Length 1089;
Best Local Similarity 55.1%; Pred. No. 9.7e-33;
Matches 351; Conservative 0; Mismatches 271; Indels 15; Gaps 2;

QY 2 CGCAGCCCTGGCAGCGCGCACTGTGTCATGAAAAAGAAATGTTCTGCTCGGGCGTCTGG 61
Db 855 CCCACCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGCGGAGCGCTCTGG 796
QY 62 TGCATCCGACAGTGGTGTCTGTCAGCCGCGACACTGTTCCAGAACTCTTACACCATCGGGC 121
Db 795 TCAATGAGCGCTGGTGTCTACTGCGCCCACTGCAAGATGAATGATACACCGTGCACC 736

QY 122 TGGCCTGCACAGTCTTGTAGAGCCGACCAAGACCAGGAGCCAGATGGTGGAGCCAGCC 181
Db 735 TGGCAGTGTATAGCTGGGCGACAGAGAG-----CTCAGAGATCAAGGCTCGA 685
QY 182 TCTCCGTACGCGACCCAGAGTACACAGACTCTTGTCTCGCTAACGACCTCATGCTATCA 241
Db 684 AGTCATTCCGCCACCCCGGCTACTCTCACACAGACCATTGTAATGACCTCATGCTGTGA 625
QY 242 AGTTGACCAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCATGCTTGGCAGT 301
Db 624 AGCTCAATAGCCAGGCCAGGCTGTCCATCCATGTTGTAAGAAAGTCAGGCTGCCCTCCGCT 565
QY 302 GCCCTACCGGGGGAACCTCTTGCTTCTGCTGCTGGGCTCTGCTGGCGAAC-----G 355
Db 564 GCGAACCCCTGGAACCACTGACTGTCTCCGCTGGGGCACTACACGAGCCAGATG 505
QY 356 GCAGATGCCCTACCGTGTCTGACATGCGGTGAACGTGTGGTGTCTGAGANGTCTGCA 415
Db 504 TGACCTTCCCTCTGACCTCATGTGGGTGATGTCAAGCTCATCTCCCCCAGACTGCA 445
QY 416 GTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTTGGCCGCGGAGGCGCAAGACC 475
Db 444 CGAAGTTTACAAGGACTTACTGGAATAATTCATGCTGTGCGCTGCGCATCCCGACTCCA 385
QY 476 AGAAGACTCCTGCAACGGTACTCTGGGGGCCCTGATCTGCAACGGGTACTTGACAG 535
Db 384 AGAAAAAGCCTGCAATGCTGACTCAGGGGACCGTTGTTGTGACAGAGTACCTGCAAG 325
QY 536 GCCTTGTGTCTTTCGAAAAAGCCCCGTGTGGCCAACCTGGCGTGGCAGGTGTCTACACCA 595
Db 324 GTCTGTTGCTCTGGGAACCTTCCCTTGGCGGCAACCAATGACCCAGAGTCTACACTC 265
QY 596 ACCTCTGCAATTCACCTGAGTGGATAGAGAAAACCGT 632
Db 264 AAGTCTGCAAGTTCACCAAGTGATTAATGACACCAT 228

RESULT 8

US-08-557-146-1
; Sequence 1, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 986 base pairs

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 25..786
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 25..90
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 112..783
: US-08-557-146-1

Query Match 13.1%; Score 153; DB 4; Length 986;
Best Local Similarity 54.9%; Pred. No. 2.5e-32;
Matches 350; Conservative 0; Mismatches 272; Indels 15; Gaps 2;

OY 2 CGCAGCCCTGGCAGGGCGCAGTGTGATGGAACGAATTGTCTGCTCGGGCTCTGG 61
DB 143 CCCACCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTCGGAGGGCTCTGG 202
OY 62 TGCATCCGCAAGTGGGTGCTGTGACCCGCGACACTGTTTCCAGAAGTCTTACACATCGGGC 121
DB 203 TCAATGAGCGCTGGGTGCTCAGTGGCCGCGCACTGCAAGATGAATGATACACCGTGACCC 262
OY 122 TGGGCTGCACAGTCTTGTAGAGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC 181
DB 263 TGGGCAGTGATACGCTGGGCGACAGGAGAG-----CTCAGAGATCAAGGCTCTCGA 313
OY 182 TCTCCGTACGGCACCAGAGTACAAAGACTCTTGTGCTTAAGACCTCATGCTCATCA 241
DB 314 AGTCAATCCGCCACCCCGGCTACTCCACACAGACCAATGTTAATGACCTCATGCTCTGA 373
OY 242 AGTTGACGAATCCGTGTCCGAGTCTGACACCAATCCGAGCATCAGCATTCGTTCCAGT 301
DB 374 AGCTCAATAGCCAGGCCAGGCTGTCAATCCATGTTGAAGAAAGTCAGGCTGCCCTCCGCT 433
OY 302 GCCCTACCGCGGGGGAACCTCTGCTCCTGCTNTCTGCTGGGCTCTGCTGGCGA-----ACG 355
DB 434 GCGAACCCTTGGAAACCACTGTACTGTCTCCGGCTGGGGCACTACCAAGAGCCAGATG 493
OY 356 GCAGATGCTTACCGTGTGCTGACATGCTGGAACGTGTGGTGTGCTGTAGAGANGTCTGCA 415
DB 494 TGACCTTTCCTCTGACCTCATGTGCGTGATGTCAAGCTCATCTCCCCCAGGACTGCA 553
OY 416 GTAAGCTATGACCCGCTGTACCAACCCAGCATGTCTGCGCCGCGAGGCAAGACC 475
DB 554 CGAAGGTTACAAGGACTTACTGTGAAAAATTCATGTGCTGTGCGCTGCGATCCCGACTCCA 613
OY 476 AGAAGGACTCTTGCACAGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGACAG 535
DB 614 AGAANAAGCGCTGCAATGTTGACTCAGGGGAGACCGTGTGTGTCAGAGGTACTCTGCAAG 673
OY 536 GCCTGTGTCTTTTGGAAAAAGCCCGTGTGGCCACTTGGCGTGCAGGTGTCTACACCA 595
DB 674 GTCTGTGTCTCTGGGGAACCTTTCCTTGGCGCAACCCCAATGACCCAGAGGTACTACTC 733
OY 596 ACCTTCGAATTCACCTGAGTGAATAGAAAAACCGT 632
DB 734 AAGTGTGCAAGTTCACCAAGTGAATAATGACACCAT 770

RESULT 9
US-08-308-949A-1/c
: Sequence 1, Application US/08308949A
: Patent No., 5580703

: GENERAL INFORMATION:
: APPLICANT: Kotin, Robert M.
: APPLICANT: Berns, Kenneth I.
: APPLICANT: Linden, Ralph M.
: TITLE OF INVENTION: Human Adeno-Associated Virus Integration
: TITLE OF INVENTION: Site DNA and Uses Thereof
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/308,949A
: FILING DATE: September 20, 1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/947,127
: FILING DATE: September 27, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Carroll, Alice O.
: REGISTRATION NUMBER: 33,542
: REFERENCE/DOCKET NUMBER: ACC92-10F
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4060 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-308-949A-1

Query Match 12.2%; Score 142.2; DB 1; Length 4060;
Best Local Similarity 71.6%; Pred. No. 3.4e-29;
Matches 240; Conservative 0; Mismatches 81; Indels 14; Gaps 4;

OY 718 CCAAGCCCTCTCTCCCTCAGGCCAGAGTCCAGGCCCCAGCCCTCTCCCTCAACCA 777
DB 4000 CCAAGCCCTCTCCCTCAACCAAGAGGCCAGGCCCCAGCCCTCTCCCTCAGATGAA 3941
OY 778 AGGTACAGATCCCAAGCCCTCTCTCCCTCAGACCAAGAGTCCAGACCCCGAGCCCT 837
DB 3940 GGAGTCCAGGCCCCCAAGCTCTCCCATTCAGACCAAGGGGTCCAG---GCCAGCCCCG 3884
OY 838 CNTCCNTCAGACCCAGAGTCCAG-----CCCTCTCCNTCAGACGACGAGTCCA 889
DB 3883 CCTCCCTAAGACCCAGAGTCCAGGCCCGCCGACCCCTCTCTCCCTCAGACCAAGAGTCCA 3824
OY 890 GACCCCCAGCCCNCTCCGTCAGACCCAGGGGTGACAGGCCCCCAACCCCTCNTCCNTC 949
DB 3823 G--GCCCCAGCCCTCTCCCTCGAGCCAGAGTCCAGGCCCGCAGTCCCTCCACCCCTC 3766
OY 950 AGAGTCAGAGTCCAGACCCCAACCCCTCTCTCCCTCCAGACCAAGAGTNCAGGTCCAG 1009
DB 3765 AGACCCAGAGTCC-AGGCCCAAGCCCTCTCTCTCGGACCCAGAGAGTCCAGGCCCCAG 3707
OY 1010 CCCCTCTCTCTCAGACCAAGCGGTCCATGCCAC 1044
DB 3706 CCCCTCTCTCTCAACCCAGAGGCCAGGCCCC 3672

RESULT 10


```
Db 190 TNGCCCCAGATGGTTCCTGACAGCAGCCCACTGCTTNAAGCCCCGCTACATAGTTCACC 249
QY 122 TGGGCTGCACAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC 181
Db 250 TGGGCGACACAACTCCAGAGGAGGAGGCTGTGAG--CAGACCCGAGACGCCACTG 306
QY 182 TCTCGTACGGCACCAGAGTACACAGACTCTTGCTCGCTAACGAC----- 230
Db 307 AGTCTTCCCCCACCACCGGCTTCAACAACAGCCTCCCAACAAGACCACCGCAATGACA 366
QY 230 TCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTGTGACACCATCCGAGCATCAGCA 289
Db 367 TCATGCTGTGAAGATGGCATGGCCAGTCTCCATCAGCTGGGCTGTGCGACCCCTCACC 426
QY 290 TTGCTTCGACAGTGCCTACCGGGGGGAAGTCTTGCTGCTGCTGCTGCTGCTGCTGCTG 349
Db 427 TCTCTTCACGCTGTGTCTACTGCTGGCACCAGCTGCTCATTTCCGGCTGGGGCAGCAGT 486
QY 350 CGAACGGC-----AGAATGCTTACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
Db 487 CCAGCCCCCAGTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
QY 404 AGGAGTCTGCAGTAACTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCCGCG 463
Db 547 ACCAGAGTGTAGAACGCTTACCCCGGCAACATCAGACACACCATGTTGCTGCTGCTG 606
QY 464 GAGGCAAGACAGAGAGACTCTGCAACGCTGACCTGAGGCGGCGCCCTGATCTGCAAG 523
Db 607 TGCAGGAGAGGGGCAAGAGACTCTGCGCAGGCTGACTCGGGGGCCCTCTGCTGTGTA 666
QY 524 GGTACTGACAGGCTGTGTCTTTGGAAGAGCCCGCTGTGCGCAACTTGCGCTGCCAG 583
Db 667 AGTCTCTCAAGCATATCTCTGCGGCGCAGATCCGTGTGCGATCACCAGAACCTG 726
QY 584 GTGCTACACCAACTCTGCAATTCAGTGAAGTGAAGAGAAACCGTCCAGNCCAGTT 643
Db 727 GTGCTACAGCAAGTCTGCAATATATGTGAGTGAAGTGAAGAGAGATGAAGAACAA 786
QY 644 A 644
Db 787 A 787
```

RESULT 12

US-08-684-862-11

Sequence 11, Application US/08684862

Patent No. 5759541

GENERAL INFORMATION:

APPLICANT: Bach, Alfred

APPLICANT: Hilten, Heinz

APPLICANT: Bialojan, Siegfried

TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and use

TITLE OF INVENTION: Thereof

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Keil & Weinkauff

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage

COMPUTER: IBM AT-compatible, 80286 processor

OPERATING SYSTEM: MS-DOS version 5.0

SOFTWARE: WordPerfect version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/684,862

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361,705

```

: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/966,040
: FILING DATE: 30-DEC-1992
: APPLICATION NUMBER: PCT/EP91/01361
: FILING DATE: 19-JUL-1991
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 957 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: MOLECULE TYPE: linear
: ORIGINAL SOURCE:
: ORGANISM: Agkistrodon rhodostoma
: LOCATION: 210 to 911
: OTHER INFORMATION: the coding region shown in (2)(ix)(B)
: OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
US-08-684-862-11
```

Query Match 8.2%; Score 95.8; DB 3; Length 957;
Best local similarity 49.9%; Pred. No. 5.3e-17;
Matches 297; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

```
QY 40 TTGTTCTGCTCGGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99
Db 279 TTGTTCTGCTCGGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
QY 100 CAGAACTCTACACCATCGGGCTGGGCTGCAAGTCTTGAGGCGGCAAGAGCCAGG 159
Db 339 AGTACGAGTCTCAAGATGAAGTTGTGATGATGACAAAGGATGAGATGAG 398
QY 160 AGCCAGATGTGAGAGCGCACTCTCCGTAGCGCACCCAGAGTACAGACACTCTGCTC 219
Db 399 CAGACAAAGAAAGCAAGAAAGTTTCATTTGCCAATAAGAAAGATGAGTACTG 458
QY 220 GCTAACGACCTCATGCTCATCAAGTGTGACGAATCCGTGTCCGAGTGTGACACCATCCGG 279
Db 459 GACAAAGACATATATGTGATCAAGCTGAACCATCTCTGTAGCAATGATGAACATCCGG 518
QY 280 AGCATCAGCATTGTCTCGAGTGCCTTACCAGCGGGGAACCTCTGCTGCTGCTGCTGCTG 339
Db 519 CCTCTCAGCTTGCCTTCCAGCCCTCCAGTGTGGGCTCATTTGCCAATATATGGATGG 578
QY 340 GGTCTGCTGGCGAAGCGGCAAGATGCTTACC-----GTGCTGCACTGCGTGAACGTGTG 393
Db 579 GGTCTCAGTACACCTGTAAAGTGAAGTCTTCCCGATGTCCCTCATTTGCTTAACATTAA 638
QY 394 GTGCTGTGAGGAGTGTGAGTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCT 450
Db 639 CTACTGATGATGACAGAGTGTGATGACAGTGTGACCTGAGTGTGAGTGTGAGTGTGAG 698
QY 451 TTCTGCGCGCGCGAGGCGCAAGACAGAAAGACTCTGCAACGCTGCTGAGGCGGCC 510
Db 699 TTGCTGACAGTATCGTGAAGAGGCAAGATATATGATGATGATGATGATGATGATGATG 758
QY 511 CTGATCTGCAACGGGTACTTGCAGAGGCGCTTGTCTTTGGAAGAAAGCCCGTGTGCCAA 570
Db 759 CTCTATCTGTAATGAACAAGTCCAGGGCATTTATCTTATGAGGCGCGCATCTGTGCCAA 818
QY 571 CTTGGCGTGCAGGTTGTCTACACCAACCTCTGCAAAATTCAGTGAAGTGAATGAGA 625
Db 819 CCTTTAAGCCTGTATCTACACCAAGGCTCCATGATTAATGATGATGATGATGATGATG 873
```

RESULT 13

US-08-358-782D-14

Sequence 14, Application US/08358782D

Patent No. 5674682

GENERAL INFORMATION:

APPLICANT: Croce, Carlo


```

1  APPLICANT:  Gomella, Leonard
2  APPLICANT:  Mulholland, S. Grant
3  APPLICANT:  Moreno, Jose
4  APPLICANT:  Fischer, Rainer
5  TITLE OF INVENTION:  Methods of Detecting Micrometastasis of Prostate
6  NUMBER OF SEQUENCES:  14
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
9  STREET:  One Liberty Place 46th. Floor
10 CITY:  Philadelphia
11 STATE:  PA
12 ZIP:  19103
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE:  Floppy disk
16 COMPUTER:  IBM PC compatible
17 OPERATING SYSTEM:  PC-DOS/MS-DOS
18 SOFTWARE:  Patentln Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER:  US/08/358,782D
21 FILING DATE:  15-DEC-1994
22
23 CLASSIFICATION:  435
24
25 ATTORNEY/AGENT INFORMATION:
26 NAME:  Beardell, Lori Y.
27 REGISTRATION NUMBER:  34,293
28 REFERENCE/DOCKET NUMBER:  TJU-1327
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE:  215-568-3100
31 TELEFAX:  215-568-3439
32
33 INFORMATION FOR SEQ ID NO:  14:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH:  1462 base pairs
36 TYPE:  nucleic acid
37 STRANDEDNESS:  single
38 TOPOLOGY:  linear
39 MOLECULE TYPE:  DNA (genomic)
40 HYPOTHETICAL:  NO
41 ANTI-SENSE:  NO
42
43 US-08-358-782D-14

```

```

Query Match      7.48; Score 85.8; DB 2; Length 1462;
Best Local Similarity 49.68; Pred. No. 3e-14;
Matches 336; Conservative 0; Mismatches 300; Indels 42; Gaps 3;

QY      2 CGCAGCCCTGGCAGCGCGCACATGTCATGGAACAACGAATTGTCTGCCTCGGCCCTCTGG 61
        ||| ||||| ||||| | ||| | | ||||| ||||| ||||| ||||| |||||
Db     128 CCCAACCCCTGGCAGGTGCTTGTGGCCTCTCGTGCGAGGGCAGCTCTGCGCGGCTGTCTGG 187

QY      62 TGCATCCGCGCAGTGGGTGCTGTCAAGCCGCACACTGTTCCAGAACTCTACACCATCGGGC 121
        ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db     188 TGCACCCCCAGTGGGTCTCTCACAGCTGCCCACTGCATCAGGAACAAAGCGTGATCTTGC 247

QY      122 TGGGCGCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGGAGCCAGATGTGTGAGGCCAGCC 181
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     248 TGGGTGGGCACAGCCTGTTTCATCTCTGAAGACACAG--GCCAGGTATTTCAGGTCAGCC 304

QY      182 TCTCCGTACGGCACCACAGAGTAACAAGACTCTTGCTCGCTA----- 224
        ||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     305 ACAGCTTCCACACACCCGCTCTACGATATGAGCCTCTGGAAGAATGCATTCTCAGGCCAG 364

QY      224 -----ACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTG 268
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     365 GTGATGACTCCAGCCACGACACCTCATGCTGCTCCGCCCTGTACAGAGCTTGCCGAGCTCACGG 424

QY      269 ACACCATCCGGAGCATCAGCATTCGCTTCGCAAGTGCCCTACCCGCGGGGAACCTTTGCCCTCG 328
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     425 ATGCTGTGAAGGTCATGGACCTGCCCAACCCAGGAGCCAGCAGCACTGGGGACCACTGCTACG 484

QY      329 TMTCTGGCGTGGG-----TCTGCTGGCGGAACCGCAGAATGCTTACCGTGTCTGCACCTGCG 382
        ||| ||||| ||||| | | | | | | | | | | | | | | | | | |
Db     485 CCTCAGGCTGGGGCAGCATTTGAACCAAGAGGAGTTCTTGACCCCAAAGAACTTCAGTGTG 544

QY      383 TGAACGTGTGGGTGTGTCTGAGGANCTCTGCAGTAAGCTCTATGACCCGCTGTACCACC 442

```

D_b 545 TGGACCTCCATGTATTATTTCCAATGACGTGTGTGCCGAAGTTCACCCTCAGAAGGTGACCA 604
QY 443 CCAGCATGTTCTTGCGCCGGCGGAGGGCAAGACCAGAAGGACTCTGC AACGGTGACTCTG 502
D_b 605 AGTTCATGCTGTGTGCTGGACGCTGGACAGGGGGCAAAGCACCTGCTCGGGTGATTTG 664
QY 503 GGGGGCCCCGTGATCTGCAACGGGTACTTGCAGGGCCTGTGTCTTTCGAAAGCCCCCT 562
D_b 665 GGGGGCCACTGTCTGTATATGTTGTCTCAAGGTATCACGTCATGCGGGCAGTGAACCAT 724
QY 563 GTGGCCCACTGGCGGTGCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGAATGATAG 622
D_b 725 GTGCCCTGCCCGAAGAGCCCTTCCCCTGTACACCAAGGTGGTGCAATTACCGAAGTGAATCA 784
QY 623 AGAAAACCGTCCAGNCCA 640
D_b 785 AGGACACCATGTTGGCCA 802

```

RESULT 14
US-08-358-782D-13
; Sequence 13, Application US/08358782D
; Patent No. 5674682
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Gomella, Leonard
; APPLICANT: Mulholland, S. Grant
; APPLICANT: Moreno, Jose
; APPLICANT: Fischer, Rainer
; TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,782D
; FILING DATE: 15-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: TJU-1327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 992 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-358-782D-13

```

```
Query Match          7.28;   Score 84.2; DB 2; Length 992;
Best Local Similarity 49.4%; Pred. No. 6.9e-14;
Matches    335; Conservative      0; Mismatches    301; Indels     42; Gaps       3;
```

QY 2 CGAGACCTGGCAGGCGGCACATGSHCATGAAACGAATTGTCTGTCGGCGTCTTG 61
 | | | | | | | | | | | | | | |
Db 147 CCCAACCTGGCAGGTGTTGTGACCCTCTCGTGGCAGGCCAGTCCGGCGGTCTTG 206
 | | | | | | | | | | | | | | |


```
QY 62 TGCATCCGAGTGGGTGCTGTACGCCGACACTGTTTCCAGAACTCCTACACCATCGGGC 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 TGCACCCCCAGTGGGTCTTCACAGCTGCCCCACTGCATCAGGAACAAAGCGTATCTTGC 266
QY 122 TGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCGAGGAGCCAGATGGTGAAGGCCAGCC 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 TGGGTGGGACAGCCTGTTTCATCCTGAAAGACACAG--GCCAGTATTTCAGGTGAGCC 323
QY 182 TCTCCGTACGGCAGCCAGAGTACAAACAGACTCTTCTCGCTA----- 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 ACAGCTTCCACACCCGCTCTACGATATGAGCCTCTGAAAGATCGATTCTCAGGCCAG 383
QY 224 -----ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTCTG 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GTGATGACTCCAGCCAGACCTCATGCTGCTCCGCTGTACAGAGCCTGCCGAGCTCAGCG 443
QY 269 ACAACATCCGGAGCATCAGCATTCGCTTGCAGTGCCTACCGGGGGAATCTTGCCTCG 328
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 ATGCTGTGAAGTCAATGACCTGCCACCCAGGAGCCAGCAGCTGGGAGCCACTGCTACG 503
QY 329 TTTCTGGCTGGGG-----TCTGCTGGCGAAGCGGCAATGCTTACCGTGTGCACTGCG 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 CCTCAGGCTGGGGCAGCATGTAACCAAGAGAGTTCTTGACCCCAAGAACTTCAGTGTG 563
QY 383 TGAAGCTGTGCTGTGTGTGAGGANGTGTGCAAGTCTATGACCCGCTGTACACC 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 TGAACCTCATGTTATTTCATGACGTGTGTGCGCAAGTTCACCCCTCAGAAAGGTGACCA 623
QY 443 CCAGCATGTTCTGCGCGGCGGAGGCAAGACCAAGAGACTCCTGCAACGCTGACTCTG 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 AGTTCATGCTGTGTGTGAGAGCTGAGACAGGGGCAAAAGACCTGCTCGGCTGATCTG 683
QY 503 GGGGCCCCCTGATCTGCAAGGGGTACTTGACAGGCGCTTGTCTTTCGAAAAGCCCCGT 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 GGGGCCCACTGTGTGTATGATGTGTCTCAAGGTATCACGTCATGGGGCAGTGAACCAT 743
QY 563 GTGCCAACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAATTCTACTAGTGGATAG 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 GTGCCCTGCCCCGAAAGGCTTCCCTGTACACCAAGGTGTGCTTACCCGGAAGTGATCA 803
QY 623 AGAAAACCGTCCAGNCCA 640
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 AGAACACCATCGTGCCA 821
```

RESULT 15
US-08-844-024-1
; Sequence 1, Application US/08844024
; Patent No. 5840494

; GENERAL INFORMATION:
; APPLICANT: Katz, Aaron E., et al.
; TITLE OF INVENTION: A Method For Molecular Staging Of
; TITLE OF INVENTION: Prostate Cancer
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,024
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,391

```
; FILING DATE: 15-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 43677/JPW/TEP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1729 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 378..1088  
; US-08-844-024-1
```

Query Match 7.2%; Score 84.2; DB 4; Length 1729;
Best Local Similarity 49.4%; Pred. No. 8.5e-14;
Matches 335; Conservative 0; Mismatches 301; Indels 42; Gaps 3;

```
QY 2 CGCAGCCCTGGCAGGCGGCACTGTGTATGGAAGAAAGCAATTGTTCTGCTGGCGCTCTGG 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 CCCAACCTGGCAGGTGCTTGTAGCCTCTGCTGCGAGGCGAGTCTGCGGCTGTTCTGG 468
QY 62 TGCATCCGAGTGGGTGCTGTACAGCCGCAACTGTTCCAGAACTCCTACACCATCGGGC 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 TGCACCCCGAGTGGGTCTTACAGCTTACAGCTGCATCAGAAACAAAGCGTATCTTGC 528
QY 122 TGGGCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGGTGAAGGCCAGCC 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TGGGTGGGACAGCGCTTTCATCCTGAAGACACAG--GCCAGTATTTCAGGTGAGCC 585
QY 182 TCTCCGTACGGCAGCCAGAGTACACAGACTCTTGTCTGCTA----- 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 586 ACAGCTTCCACACACCCGCTCTACGATATGACCCCTCTGAAGATCGATTCTCAGGCCAG 645
QY 224 -----ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTCTG 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 GTGATGACTCCAGCCAGACCTCATGCTGCTCGGCTGTACAGAGCCTGCCGAGCTCAGCG 705
QY 269 ACACCATCCGGAGCATCAGCATTCGCTTGCAGTGCCTTACCGGGGGAACCTCTGCTCG 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 ATGCTATGAAGTCAATGAGCTGCCCCACCCAGGAGCCAGCAGCAGTGGGAGCCACTGTA 765
QY 329 TTTCTGGCTGGGG-----TCTGCTGGCGAAGCGGCAAGATGCTTACCGTGTGCACTGCG 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 CCTCAGGCTGGGCGACATTTGAACCAAGAGAGTCTTGACCCCAAGAAACTTCAGTGTG 825
QY 383 TGAAGCTGTGCTGTGTGTGAGGANGTGTGCAAGTAACTCTATGACCCGCTGATACCAC 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 TGAACCTCATGTTATTTCATGACGTGTGTGCGCAAGTTCACCTCAGAAAGGTGACCA 885
QY 443 CCAGCATGTTCTGCGCGGCGGAGGAGGCAAGACCAAGAGACTCTGCAACGGTGAATCTG 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 AGTTCATGCTGTGTGTGAGCGCTGAGACAGGGGGGCAAAAGACACTGCTCGGCTGATTC 945
QY 503 GGGGCCCCCTGATCTGCAACGGGTACTTGACAGGGGCTTGTGCTTTCGAAAAGCCCCGT 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 GGGGCCCACTGTGTGTATGATGTGTGCTTCAAGGTATCACGTCATGAGGGGAGTGAAC 1005
QY 563 GTGCCAACTTGGCGTGCAGAGTGTCTACACCAACCTCTGCAATTCTACTGAGTGGATAG 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 GTGCCCTGCCCCGAAAGGCTTCCCTGTACACCAAGGTGTGATTAACCGGAAGTGATCA 1065
QY 623 AGAAAACCGTCCAGNCCA 640
```


Db 1066 AGGACACCATCGTGGCCA 1083

Search completed: September 28, 1999, 11:34:10
Job time: 2130 sec

This Page Blank (uspto)

OM of: US-09-030-606-175 to: PIR_60:* out_format : pfs
Date: Sep 25, 1999 6:28 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171617_29825/app-query.fasta.1
-DB=PIR_60 -QFMT=fastan -SUFFIX=xpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-030-606-175
Query length: 1167
Database: PIR_60:*
Database sequences: 122810
Database length: 40065486
Search time (sec): 182.540000

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|-------------|-----------|--------|---------|-----|----------------------------------|
| PIR2:A53968 | 530.50 | 561.04 | 6.1e-24 | 253 | serine proteinase SCCE precursor |
| PIR2:I56559 | 498.50 | 527.55 | 4.3e-22 | 260 | neutropsin - mouse |
| PIR2:A44284 | 478.00 | 506.68 | 6.7e-21 | 244 | tissue kallikrein (EC 3.4.21.35 |
| PIR1:NGMSG | 472.00 | 499.96 | 1.5e-20 | 261 | 7S nerve growth factor gamma ch |
| PIR2:S55065 | 470.50 | 498.76 | 1.8e-20 | 248 | trypsin (EC 3.4.21.4) I precurs |
| PIR2:A27547 | 469.50 | 497.75 | 2.1e-20 | 247 | trypsin (EC 3.4.21.4) precursor |
| PIR2:S55067 | 468.50 | 496.68 | 2.4e-20 | 248 | trypsin (EC 3.4.21.4) I precurs |
| PIR2:S55066 | 467.50 | 495.64 | 2.7e-20 | 248 | trypsin (EC 3.4.21.4) II precu |
| PIR2:S45303 | 466.00 | 493.71 | 3.3e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A34079 | 463.00 | 490.59 | 4.9e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:B31136 | 463.00 | 490.65 | 4.9e-20 | 259 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:JE0236 | 462.00 | 489.55 | 5.6e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:JQ1472 | 462.00 | 489.98 | 5.7e-20 | 246 | trypsin (EC 3.4.21.4) V precurs |
| PIR2:S33772 | 461.50 | 489.14 | 6.0e-20 | 257 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A31136 | 461.00 | 488.51 | 6.4e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:S01971 | 461.00 | 488.51 | 6.4e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR1:TRDGC | 459.50 | 487.38 | 7.9e-20 | 246 | trypsin (EC 3.4.21.4) precursor |
| PIR2:JQ1471 | 459.00 | 486.86 | 8.4e-20 | 246 | trypsin (EC 3.4.21.4) V precurs |
| PIR2:A41020 | 458.00 | 485.39 | 9.6e-20 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A29746 | 458.00 | 485.45 | 9.6e-20 | 259 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:S05494 | 457.00 | 484.75 | 1.1e-19 | 247 | trypsin (EC 3.4.21.4) IV precu |
| PIR1:TRBOTR | 456.50 | 484.77 | 1.2e-19 | 229 | trypsin (EC 3.4.21.4) precursor |
| PIR2:A29745 | 452.00 | 479.15 | 2.1e-19 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR1:EGMSB | 450.00 | 477.07 | 2.8e-19 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A29586 | 449.00 | 476.03 | 3.2e-19 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A32297 | 448.50 | 475.93 | 3.4e-19 | 246 | prostate-specific serine prote |
| PIR2:B25528 | 448.00 | 475.93 | 3.4e-19 | 246 | trypsin (EC 3.4.21.4) precursor |
| PIR1:KORTN | 448.00 | 475.04 | 3.6e-19 | 259 | tonin (EC 3.4.21.-) precursor |
| PIR1:KORTP | 447.00 | 473.84 | 4.2e-19 | 265 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A25606 | 446.00 | 472.91 | 4.8e-19 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:A35871 | 445.00 | 472.38 | 5.5e-19 | 243 | trypsin (EC 3.4.21.4) precursor |
| PIR2:T01779 | 443.50 | 470.61 | 6.7e-19 | 250 | trypsin (EC 3.4.21.4) - plai |
| PIR1:TRRT1 | 442.50 | 469.69 | 7.6e-19 | 246 | trypsin (EC 3.4.21.4) I precurs |
| PIR2:S31779 | 442.00 | 469.93 | 7.7e-19 | 238 | trypsin (EC 3.4.21.4) III precu |
| PIR1:TRMSM5 | 442.00 | 468.74 | 8.1e-19 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:S13813 | 441.50 | 468.62 | 8.7e-19 | 247 | trypsin (EC 3.4.21.4) - bovine |
| PIR1:TRPCTR | 440.50 | 468.06 | 1.0e-18 | 231 | trypsin (EC 3.4.21.4) precursor |
| PIR1:TRDG | 438.50 | 465.50 | 1.3e-18 | 247 | trypsin (EC 3.4.21.4) precursor |
| PIR1:TRDFS | 438.00 | 465.52 | 1.4e-18 | 229 | trypsin (EC 3.4.21.4) precursor |
| PIR1:KQMS1 | 437.00 | 463.54 | 1.6e-18 | 261 | tissue kallikrein (EC 3.4.21.35 |
| PIR1:KQHU | 436.50 | 462.99 | 1.7e-18 | 262 | tissue kallikrein (EC 3.4.21.35 |
| PIR2:S35711 | 435.00 | 461.46 | 2.1e-18 | 261 | tissue kallikrein (EC 3.4.21.35 |

PIR1:A25852 + 434.50 461.34 2.2e-18 247 I trypsin (EC 3.4.21.4) I prec
PIR2:S31778 + 434.00 461.30 2.4e-18 231 I trypsin (EC 3.4.21.4) II pre
PIR2:S49489 + 430.00 456.80 4.1e-18 242 I trypsin (EC 3.4.21.4) precu
seq_name: PIR2:A53968

seq_documentation_block:
serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 24-Sep-1998
C:Accession: A53968
R:Hansson, L.; Stromqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egeldrud, J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en
A:Reference number: A53968
A:Accession: A53968
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:g521214; PID:g532504
C:Genetics:
A:Gene: GDB:PRSS6; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRY>

alignment_scores:
Quality: 530.50 Length: 214
Ratio: 3.255 Gaps: 3
Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:

US-09-030-606-175 x A53968 ..
Align seg 1/1 to: A53968 from: 1 to: 253

```
1 GCGCAGCCCTGGCAGCGCGCAGCTGTCATGAAACGAATTGTTCTGTC 50
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG1 56
   56 yGlyValLeuValAsnGlnValArgTrpValLeuThrAlaAlaHisCysLysM 73
101 AGACTTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGCCGACCA 150
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
73 etAsnGluTrpThrValHisLeuGlySerAspThrLeu..GlyAspArg 88
151 GAGCCAGGGAGCCAGATGTGGAGGCCAGCCTTCCTGACGGCACCAGA 200
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG1 103
201 GTACACAGACTCTTGCTGCTGCTAACGACCTCATGCTCAAGTTGAGC 250
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
103 YTrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsn 120
251 AATCCGTCGTCGAGTCTGACACCATCCGGAGCATGACATGCTTCGCAG 300
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
301 TGCCCTACCCGGGGAGACTCTTGCCCTCGTNTCTGCTGGGCTGCTGCG 350
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
137 CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153
351 GAACGCG.....AGAATGCTACCGTGTGCTGCTGCTGCTGACGCTGCG 394
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170
395 TGGTGTCTGAGGANGTCTGACGTATGACCTGATGACCCGCTGTACACCCC 444
   :::::::::::::::::::::  :::::::::::::::::::::  :::::::::::::::::::::
```


170 eulieserProglInaspCysThrLysValTyrLysAspLeuGlnAsn 186
445 AGCATGTTCTGCGCGCGGAGGCAAGACCAAGAACTCCTGCACGG 494
187 SermetLeuCysAlaGlyIleProAspSerLysAsnAlaCysAsnG1 203
495 TGACTCTGGGGGCCCCCTGATCTGCAACGGGTACTTCAGGGCCTGTGT 544
203 yaspserGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuValS 220
545 CTTTCGGAAGAGCCCCGTGTGCGCAACTGGCGTGCACAGGTCTACACC 594
220 erTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236
595 AACCTCTGCAAAATTCACGTGAGTGATAGAGAAAACCGTCAG 636
237 GlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 250
seq_name: plir2:I56559

seq_documentation_block:
neuropsin - mouse
C/Species: Mus musculus (house mouse)
C/Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 08-Sep-1997
C/Accession: I56559
R/Chen, Z.-L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nishin
J. Neurosci. 15, 5088-5097, 1995
A/Title: Expression and activity-dependent changes of a novel limbic-serine protease gen
A/Reference number: I56559; MUID:95348817
A/Accession: I56559
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-260 <RES>
A/Cross-references: GB:D30785; NID:g1020090; PID:g1020091
C/Superfamily: trypsin; trypsin homology
F;33-252/Domain: trypsin homology <TRY>

alignment_scores:
Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

alignment_block:
US-09-030-606-175 x I56559 ..

Align seg 1/1 to: I56559 from: 1 to: 260

1 GCGCAGCCCTGGCAGCGCGCACTGGTCATGGAACGAATTGTTCTGCTC 50
43 SerGlnProTrpGlnAlaAlaLeuPheGlnGlyIleGlyLeuIleCysG1 59
51 GGGCGTCTGTGTCATCCGAGTGGGTGCTGTACAGCCGACACTGTTTCC 100
59 yGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysLysL 76
101 AGAACTCTACACCATCGGCGCTGGCCCTGCACAGTCTTGAGCGGACCAA 150
76 ysglnLysTyrSerValArgLeuGlyAspHisSerLeuGlnSerArgasp 92
151 GAGCCAGGAGCCAGATGTTGAGAGGCCAGCCTTCCTCCGTACGGCACCAGA 200
93 GlnPro...GlnGlnGlnIleGlnValAlaGlnSerIleGlnHisProcy 108
201 GTACAACAGACTCTTG.....CTGCTAACGACTCATGCTCATCA 241
108 sTyraSnaSnaSerAsnProGlnAspHisSerHisAspIleMetLeuIleA 125
242 AGTTGAGCAATCCGTTGTCGAGTCTGACACCAATCCGAGAGCATCAGCAT 291
125 tgleuGlnAsnSerAlaAsnLeuGlyAspLysValLysProValGlnLeu 141
292 GCTTCGACAGTGCCCTACCGCGGGGAAACTCTTGCTGCTGCTGCTGGG 341

142 AlaAsnLeuCysProLysValGlyGlnLysCysIleIleSerGlyTrpG1 158
342 TCTGCTGGCGAAGCGCAGA.....ATGCCTACCGTGTGCTGCTGCTGTA 385
158 yThrValThrSerProGlnGlnAsnPheProAsnThrLeuAsnCysAlaG 175
386 ACGTGTCGTGTGTCGTGAGGANGTGTGACGTCTATGACCCGCTG 435
175 luValLysIleTyrSerGlnAsnLysCysGluArgAlaTyrProGlyLys 191
436 TACCACCCAGCATGTCTTCGCGCGCGGAGGCAAGACCAAGACTC 485
192 IleThrGlnGlyMetValCysAla...GlySerSerAsnGlyAlaAspTh 207
486 CTGCAACGGTGACTCTGGGGGCCCCCTGATCTGCAACGGGTACTTCAGG 535
207 rCysGlnGlyAspSerGlyGlyProLeuValCysAspGlyMetLeuGlnG 224
536 GCCTTGTCCTTTCGGAAGAGCCCCGTGTGCGCAACTGGCGTCCAGGT 585
224 lYleThrSerTrpGlySerAspProCysGlyLysProGlnLysProGly 240
586 GTCTACACCAACTCTGCAAAATTCACGTGATGATAGAGAAAACCGTC 633
241 ValTyrThrLysIleCysArgTyrThrThrTrpIleLysLysThrMet 256
seq_name: plir2:A44284

seq_documentation_block:
tissue kallikrein (EC 3.4.21.35) homolog - rat (fragment)
N/Alternate names: proteinase B light chain
C/Species: Rattus norvegicus (Norway rat)
C/Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C/Accession: A44284; C41429
R/Ma, J.X.; Chao, J.; Chao, L.
Biochemistry 31, 10922-10928, 1992
A/Title: Molecular cloning and characterization of rKlk10, a cDNA encoding T-kininogen
A/Reference number: A44284; MUID:93041794
A/Accession: A44284
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-244 <MA1>
A/Cross-references: GB:S48142; NID:g259430; PID:g259431
A/Experimental source: submandibular gland
A/Note: sequence extracted from NCBI backbone (NCBIN:118095, NCBIPI:118096)
R/Kato, H.; Nakanishi, E.; Enjoji, K.; Hayashi, I.; Oh-Ishi, S.; Iwanaga, S.
J. Biochem. 102, 1389-1404, 1987
A/Title: Characterization of serine proteinases isolated from rat submaxillary gland:
A/Reference number: A41429; MUID:88198057
A/Accession: C41429
A/Status: preliminary
A/Molecule type: protein
A/Residues: 10-61 <KAT>
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; serine proteinase
F;10-236/Domain: trypsin homology <TRY>
F;48,103,196/Active site: His, Asp, Ser #status predicted

alignment_scores:
Quality: 478.00 Length: 225
Ratio: 2.987 Gaps: 4
Percent Similarity: 71.111 Percent Identity: 41.333

alignment_block:
US-09-030-606-175 x A44284 ..

Align seg 1/1 to: A44284 from: 1 to: 244

1 GCGCAGCCCTGGCAGCGCGCACTGTGATGGAACGAATGTTCTGCTC 50
20 SerGlnProTrpGlnValAlaIleIle.....AsnGluTyrLeuCysG1 34


```

445 AGCATGTCTTGCGCCGCGGAGGCGCAAGACGAAAGACTCTTGCACCG 494
      ::|||::||| ||| :::: ||||| |||||:::|
184 Asnmetilecysileglytyrleuasnglyglylysaspsercysglnql 200
      ::|||::||| ||| :::: ||||| |||||:::|
495 TGACTCTGGGGGGCCCCCTGATCTGCACAGGGTACTTGCAAGCCCTGTGT 544
      ||||| ||||| |||||::: ||||| |||||:::|
200 yaspserglyglyprovalvalcysasnnglyglnleuglnglyphevals 217
      ||||| ||||| |||||::: ||||| |||||:::|
545 CTTCGGAAAGCCCCGTGTGGCCCACTTGCGGTGCCAGTGTCTACACC 594
      |||::||| ||| |||::||| ||| ||||| |||||
217 erttrpgly...ileglycysalaglnlysglytyrproglyvaltyrthr 232
      |||::||| ||| |||::||| ||| ||||| |||||
595 AACCTCTGCAAAATTCAGTGATGTGATAGAGAAAACCGTCCAGNCCAGT 642
      ::::: |||::: |||::: |||::: |||::: |||:::
233 lysvalcysasnlyrvalsertrpilelysthrthrmetseraserasn 248

```

```

seq_documentation_block:
  trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29) - chicken
N:Alternate names: trypsinogen II
C:Species: Gallus gallus (chicken)
C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 08-Sep-1997
C:Accession: S55066; S72347
R:Wang, K.; Gan, L.; Lee, I.; Hood, L.
Biochem. J. 307, 471-479, 1995
A:Title: Isolation and characterization of the chicken trypsinogen gene family
A:Reference number: S55065
A:Accession: S55066
A:Molecule type: mRNA
A:Residues: 1-248 <WAN1>
A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907
A:Experimental source: clone 2-P29
A:Accession: S72347
A:Molecule type: DNA
A:Residues: 1-248 <WAN2>
A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907
A:Experimental source: clone 2-P29
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine protease; zymogen
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-25/Domain: activation peptide #status predicted <APT>
F:26-248/Product: trypsin II #status predicted <MAT>
F:26-241/Domain: trypsin homology <TRY>
F:65,109,202/Active site: His, Asp, Ser #status predicted

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 467.50 | Length: 211 |
| Ratio: | 2.959 | Gaps: 5 |
| Percent Similarity: | 74.882 | Percent Identity: 42.654 |

```
alignment_block:
US-09-030-606-175 x S55066 .
```

Align seg 1/1 to: S55066 from: 1 to: 248

```

7  CCCTGGCAGCGGCGGCACTGGTCATCGAAGAAACGAATGTTGTTCTGCTGGGCGGT 56
   |||::|||:::||||| ::::: |||||:::|||||
38  PRTYRGLNVALSERLEU...ASNSERGITYRHNISPHECYSGLYGLYSE 53
   |||::|||:::||||| ::::: |||||:::|||||
57  CCTGGTCATCCGCGAGTGGGTGCTGTACGCCGACACTGTTTCCAGAACT 106
   |||:::||||| ||||| ||||| ||||| ||||| ||||| :::::
53  rleuileasnserglntprvalleuseralalaiahiscystyltyssera 70
   |||:::||||| ||||| ||||| ||||| ||||| ||||| :::::
107 CCTACACCATCGGGCTGGGCGCTGCACAGTCTTGAGGCCGACCAAGACCA 156
   ::: ||||| ::::: ||||| ||||| |||||
70  rgileglnvalargleuglyglutyrasnileaspval..GlnGluasp 85
   ::: ||||| ::::: ||||| ||||| |||||
157 GGGAGCCAGATGGTGGAGGCCACGCTCTCCGTACGGCACCACCCAGAGTACAA 206
   ::::: ::: ::::: ||||| ::: ||||| ||||| ||||| :::::
86  SerGluValValArgSerSerSerValIleIleLeaRghHisProLysTyrSe 102

```

207 CAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG 256
::: ||| |||||:::||||| ||||| ::::
102 rserIlethrLeuAsnAsnPileMetleuIleLysLeuAlaSerAlay 119
257 TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTCCTTCGAGTGCCT 306
||::: ||| ||::: ||::: ||::: ||
119 alGIuTyrserAlaaspIleGlInProIlealaleuProserSerCysAla 135
307 ACCGGGGGAACTCTTGCCCTCGINNTCTGGCTGGGT...CTGCTGGCGAA 353
|||||:::|||||:::||||| ||||| ||::: ||
136 LysAlaGIyThrGIuCysLeuIleSerGIyTrpGIyAsnThrLeuSerAs 152
354 CGGC..AGAATGCCCTACCGGTGCTGCACCTGCCTGAACGTGTGGTGTGT 400
||| :::: ||| ::|||:::||::: ::::
152 nGIyTyraSnTyPrProGIuLeuGIuGIuCysLeuAsnAlaProIleuS 169
401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCCGCTGACCAACCCAGCATG 450
||::: ||::: ||::: ||| ::|||
169 eraSpGIuGIuCysGIuGIuAlaTyPrProGIyAspIleThrserAsnMet 185
451 TTCTGCGCGCGGAGGGGCAAGACCAAGAGACTCTTGCAACGGTGACTC 500
::|||:::||| :::: ||||| |||||:::||||| |
186 IleCysValGIyPheLeuGIuGIyGIyLysAspSerCysGIuGIyAspSe 202
501 TGGGGGGCCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTCCTTTCG 550
||||| ||::: ||||| ||||| |||||:::|||||:::
202 relGIyGIyProValValCysasnGIyGIuLeuGIuGIyIleValserTrpG 219
551 GAAAAGCCCCGTGTGGCCAACCTTGCGCTGCCAGGTGCTTACACCAACCTC 600
|| ||::: ||| ||||| ||||| |||||:::
219 Iy..IleGIyCysAlaLeuLysGIyTyPrProGIyValTyThrIlyVal 234
601 TGCAATTCACCTGAGTGGATAGAAAAACCGTC 633
|||::: ||::: |||||:::|||||:::||:::
235 CysasnTyrValaspTrpIleGIuGIuThrIle 245

```

seq_documentation_block:
tissue kallikrein (EC 3.4.21.35) precursor - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 20-Oct-1994 #sequence_revision 03-Aug-1995 #text_chan
C/Accession: S45303; S38487
R/Gauthier, E.R.; Dumas, C.; Chapdelain, P.; Tremblay, R.R.;
Biochim. Biophys. Acta 1218, 102-104, 1994
A/Title: Characterization of canine pancreas kallikrein cDNA.
A/Reference number: S45303; MUID:94250683
A/Accession: S45303
A/Molecule type: mRNA
A/Residues: 1-261 <GAU>
A/Cross-references: EMBL:X75479; NID:g414018; PID:g414019
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; serine proteinase
F/1-24/Domain: signal sequence #status predicted <SIG>
F/25-261/Product: tissue kallikrein #status predicted <MAT>
F/25-253/Domain: trypsin homology <TRY>
F/65,120,213/Active site: His, Asp, Ser #status predicted

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 466.00 | Length: 224 |
| Ratio: | 2.913 | Gaps: 3 |
| Percent Similarity: | 71.429 | Percent Identity: 39.732 |

```
alignment_block:
US-09-030-606-175 x S45303 .
```

Align seg 1/1 to: S45303 from: 1 to: 261

1 GCGCAGCCCTGGCAGCGGCACATGGTCATGAAACGAATTGTTCTGCNC 50
:::|||||
35 SerGlnProTrrpGlnAlaAlaLeuTyrHisTyrSerLysPheGlnCysGcl 51


```

51 GGGCGCTCTGGTGCATCCGGCAGTGGGTGCTGCAGCCGACATGTTTCC 100
   :|||||
51 YGLYALLEUVALHISPROGLUTRVALTHRALAALAHISCYSILEA 68
   :|||||
101 AGAATCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGGACCAA 150
   :|||
68 snAspAsnTYrGlnLeuTYrPLeuGlyArgTYrAsnLeu..PheGlnHis 83
   :|||
151 GAGCCAGGGAGCCAGATGGTGAAGGCCAGCCTCTCCGTACGGACCCAGA 200
   :|||
84 GILAspThrAlaGlnPheValGlnValArgGluSerPheProHisProGl 100
   :|||
201 GTACAAACAGACTCTTGCTCGCTAC..... 225
   :|||
100 uPheAsnLeuSerLeuLeuLysAsnHisThrArgLeuProGlnGluAspT 117
   :|||
226 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCT 267
   :|||
117 YrSerHisAspIleMetLeuValArgLeuAlaGluProAlaGlnIleThr 133
   :|||
268 GACACCATCCGGAGCATCAGCATGCTTGGCAGTCCCTACCGCGGGGAA 317
   :|||
134 AspAlaValArgValLeuAspLeuProThrGlnGluProGlnValGlyse 150
   :|||
318 CTCTTGCCCTGNTCTGGCTGGGCTGCTGCGGACGGCAGAAATG... 363
   :|||
150 rThrCysTYrAlaSerGlyTYrGlySerIleGluProAspLysPheIleT 167
   :|||
364 ..CCTACCGTGCCTGCACCTCCGTGAACGTGCGGTGTCTTAGGANGTC 411
   :|||
167 YrProAspAspLeuGlnCysValAspLeuGluLeuLeuSerAsnAspIle 183
   :|||
412 TGCAGTAAGCTCTATGACCCCGCTGTACCACCCAGCATGTTCTGCGCGG 461
   :|||
184 CysAlaAsnAlaHisSerGlnLysValThrGluPheMetLeuCysAlaGl 200
   :|||
462 CGGAGGGCAAGACCCAGAGAGACTCTGCAACGGGTGACTCTGGGGGGCCC 511
   :|||
200 YHisLeuGlnGlyGlyLysAspThrCysValGlyAspSerGlyGlyProL 217
   :|||
512 TGATCTGCAACGGGTACTTGCAGGGCCTTGTTCTTTCCGAAAGCCCCG 561
   :|||
217 euIleCysAspGlyValLeuGlnGlyIleThrSerTYrGlyHisValPro 233
   :|||
562 TGTGGCCAACCTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAAATTCAC 611
   :|||
234 CysGlySerProAsnMetProAlaValTYrThrLysValIleSerHisLe 250
   :|||
612 TGAAGTGGATAGAGAAAACCGTC 633
   :|||
250 uGluTYrPLeuLysGluTYrMet 257

```

seq_name: plr2:A34079

```
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) P1 precursor - rat
  N|Alternate names: kallikrein-related proteinase k8
  C|Species: Rattus norvegicus (Norway rat)
  C|Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 08-Sep-1997
  C|Accession: A34079; S10700
  R|Brady, J.M.; Wines, D.R.; MacDonald, R.J.
  Biochemistry 28, 5203-5210, 1989
  A|Title: Expression of two kallikrein gene family members in the rat prostate.
  A|Reference number: A34079; MUID:89352606
```

A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-261 <BRA>
A:Cross-references: GB:M27215; GB:M27216; GB:M27217; NID:g206638; PID:g206640; GB:M27218
A:Experimental source: prostate
R:Elmoujahed, A.; Gutman, N.; Brillard, M.; Gauthier, F.
FEBS Lett. 265, 137-140, 1990

A/Title: Substrate specificity of two kallikrein family gene products isolated from t
A/Reference number: S10698; MUID:90306305
A/Accession: S10700
A/Molecule type: protein
A/Residues: 25-43;112-138 <ELM>
A/Experimental source: submaxillary gland
A/Note: 125-Lys was also found
C/Superfamily: trypsin; trypsin homology
C/Keywords: hydrolase; serine proteinase
F;25-253/Domain: trypsin homology <TRY>
F;25-111/Product: tissue kallikrein PI light chain #status experimental <MAT1>
F;112-261/Product: tissue kallikrein PI heavy chain #status experimental <MAT2>
F;65,120,213/Active site: His, Asp, Ser #status predicted

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 463.00 | Length: 225 |
| Ratio: | 2.987 | Gaps: 3 |
| Percent Similarity: | 68.889 | Percent Identity: 40.889 |

alignment_block:

US-09-030-606-175 x A34079

Align seg 1/1 to: A34079 from: 1 to: 261

1 GCCGAGCCCTGGCAGCGCCGCACTGTCATGGAACGAATTGTTCTGCTC 50
 :::::::::::::::::::: ||||| :::::
 35 SerGlnProTrpGlnValAlaValTyrHisPheAsnGluProGlnCysG1 51

```

51 GGGCGTCCTGGTGACAGCCAGTGGGTGCTGCAGCCGACACTGTTCC 100
   :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
51 yglYValLeuIleHisProSerTrpValIleThrAlaAlaHisCysTyr 68
```

```

101 AGAAGCTCCCTACACCATCGGGCTGGGCGCTGCACAGTCTTTGAGGCCGACCA 150
    :: ::||| :: ||||| ::::||| ||
68 erValasntyrGlnValTrrpleuGlyArgAsnAsnleuLeuGluAsp... 83

```

```

151 GAGCCAGGGAGCCAGATGTTGTGAGGCCAGCCTCTCCGTACGGCACCACAGA 200
      |||||  ::::|  |||  |||||
84 GluProPhealaGlnHisArgLeuValSerGlnSerPheProHisProG 100

```

201 GTACACACAGACTCTTGCTC..... 219
 ::||| :::::
 100 yPheasnleuaspllellelysasnhstThrArgLysProGlyAsnapt 117

```

220 ..GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGCT 267
      ::::::::::::::: || ::::: :::
117 yfSerAsnaspMetLeuLeuHisLeuLysThrProIaaspIleThr 133

```

268 GACACCATCCGGAGCATCAGCATTCGTCGACAGTGCCTACCGGGGGA 317
 || ::::: ||::: || :::::
 134 AspGlyValValIleAspLeuProThrGluGluProLysValGlySe 150

```

318 CTCATGCCCTGTCNTCTGGCTGGGTCTGCTGGCG...AACGGCAGAA 361
    ::::|::::::::::|:::::   :::::      ::
150 rThrcysleuFrnSergIyTIpqIsSerIleThrProeulystItpslup 167

```

```

362 TGCCTACCGTGCTGCACTGCGGTGAACGTTGCGTGGTGTCTGAGGANGTC 411
::||| |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
167 heProaspaspleuglncysValasnileHisleudeuseranngluls 183
```

412 TGCAGTAAGCTCTATGACCCGGCTGTACCCACCCAGCATGTTCTGCGCCGG 461
 184 CysllelysalatyrasnaspclvalThraspvalmetleucysalagl 200

462 CCGAGGGCAACAGCAGAGGACTCTCCAAACGGTACTCTGGGGGGCCCC 511
+ :: ||||| ||:::|||||
200 yclumetaspclyclyllysaspliccyslsglyspserglyglyprol 217

```

512 TGATCTGCACACGGGTACTTGCAGGGCCTTGTCGTCCTTCGGAAAAAGCCCCG 561
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
217 eutleCysaspIyValleuGlnClyIleThrSerTrypIySermetPro 233

```



```
201 GTACACAGACTCTTGCTC..... 219
      |||||
100 YTyraSmetSerLeuLeuMetLeuGlnThrIleProProGlyAlaAsp 117
220 ..GCTAACGACCTCATGCTCATCAAGTGGACGAATCCGCTCCGAGTCT 267
      :::::::::::::::::::::
117 heSerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIleThr 133
268 GACACCATCCGGAGCATCAGCATTTGCTTCGACGTGCCCTACCGCGGGAA 317
      ||::::::::: ||::::: |||
134 AspValValLysProIleAlaLeuProThrLysGluProLysProGlySe 150
318 CTCTTGCCCTCGTNTCTGGCTGGGCTGCTGCGACGACGAGA.....A 361
      :::::::::::::::::::::
150 rLysCysLeuAlaSerGlyTrpGlySerIleThrProThrArgTrpGlnL 167
362 TGCCCTACCGTGTGCACTGCGTGACGCTGCGGTGGTGTCTGAGANGTC 411
      |||
167 ysProAspAspLeuGlnCysValPheIleThrLeuLeuProAsnGluAsn 183
412 TGCAGTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTGCCGG 461
      ||::::::::: |||
184 CysAlaLysValTyrLeuGlnLysValThrAspValMetLeuCysAlaGl 200
462 CGGAGGGCAAGACCAAGAGAGACTCTGCAACGGGTGACTGTGGGGGGCCC 511
      |
200 yGluMetGlyGlyGlyLysAspThrCysArgAspAspSerGlyGlyProL 217
512 TGATCTGCAACGGGTACTTGACAGGGCTGTGCTTTGCGAAAGCCCCG 561
      ||::::: |||
217 euIleCysAspGlyIleLeuGlnGlyThrThrSerTyrGlyProThrPro 233
562 TGTGGCCAACCTGGCGGTGCCAGGTGCTTACACCAACCTCTGCAATTCAC 611
      ||::::: ||::::: ||::::: ||::::: ||:::::
234 CysGlyLysProGlyValProAlaIleTyrThrAsnLeuIleLysPheAs 250
612 TGAGTGGATAGAGAAACCGTC 633
      ::::::::::: ||:::
250 nSerTrpIleLysAspThrMet 257
seq_name: plr2:J01472
```

```
seq_documentation_block:
  trypsin (EC 3.4.21.4) V precursor, b-form - rat
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
  C:Accession: J01472
  R:Kang, J.; Wiegand, U.; Mueller-Hill, B.
  Gene 110, 181-187, 1992
  A:Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
  A:Reference number: J01471; MUID:92165057
  A:Accession: J01472
  A:Molecule type: mRNA
  A:Residues: 1-246 <KAN>
  A:Cross-references: EMBL:X59013; NID:957414; PID:957415
  A:Experimental source: pancreas
  C:Superfamily: trypsin; trypsin homology
  C:Keywords: hydrolase; protein digestion; serine proteinase; zymogen
  F:1-15/Domain: signal sequence #status predicted <SIG>
  F:16-24/Domain: activation peptide #status predicted <ACT>
  F:25-246/Product: trypsin V, b-form #status predicted <MAT>
  F:25-239/Domain: trypsin homology <TRY>
  F:31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted
  F:64,108,200/Active site: His, Asp, Ser #status predicted
```

```
alignment_scores:
  Quality: 462.00      Length: 210
  Ratio: 3.000      Gaps: 4
  Percent Similarity: 73.333      Percent Identity: 42.381
alignment_block:
```

```
US-09-030-606-175 x J01472 ..
Align seg 1/1 to: J01472 from: 1 to: 246
7 CCCTGGCAGCGGCACACTGGTCATGGAAGAATGTCTCTCGGGCGT 56
  ||:::||||:::||||| ::::: ||:::||||
37 ProTyrGlnValSerLeuAsnAlaGlySerHisIle...CysGlyGlySe 52
57 CCTGGTCATCCGACAGTGGGTCTGTACGCCGACACTGTTCCAGACT 106
  ||::: ||::::: ||::::: ||::::: ||::::: ||:::::
52 rLeuIleThrAspGlnTrpValLeuSerAlaAlaHisCysTyrHisProg 69
107 CCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCA 156
  :: :::: ||||| ||::: ||::: |||
69 InLeuGlnValArgLeuGlyGluHisAsnIle...TyrGluIleGly 84
157 GGGAGCCAGATGTTGAGGCCAGCCTCTCCGTACGCGACCCAGAGTACA 206
  :::::||||:::||||::: ||::: ||::::: ||:::
85 AlaGluGlnPheIleAspAlaAlaLysMetIleLeuHisProAspTyrAs 101
207 CAGACTCTTGCTCGCTAACGACCTCATGCTCATGAGTTGACGAATCCG 256
  :::: :::: ||::::: ||::::: ||::::: ||::::: ||:::
101 pLysTrpThrValAspAsnAspIleMetLeuIleLysSerProA 118
257 TGTCCGACTCTGACACCATCCGAGCATCAGCATGCTTCGACGTGCCCT 306
  ::::: ::::: ::::: ||::: ||::: ||::: ||::: ||:::
118 IatThrLeuAsnSerLysValSerThrIleProLeuProGlnTyrCysPro 134
307 ACCGGGGGAACCTTGCTCGCTGNTCTGGCTGGGCTGCTGGCGAACGG 356
  ||::::: ||::::: ||::::: ||::::: ||::::: ||::::: ||
135 ThrAlaGlyThrGluCysLeuValSerGlyTrpGlyValLeuLysPheGl 151
357 C..AGAATGCCCTACCGTGTGCTGCACTGCGTGAACGTGTCGGTGTCTG 403
  | :::: ||::: ||::::: ||::::: ||::::: ||::::: ||:::
151 yPheGluSerProSerValLeuGlnCysLeuAspAlaProValLeuSerA 168
404 AGGANGTCTGCACTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTC 453
  ::::: ||||| ||| ||| ::::: |||||
168 spSerValCysHisLysAlaTyrProArgGlnIleThrAsnAsnMetPhe 184
454 TCGCGCGCGGAGGGCAAGACCAAGAGACTCTGCAACGGGTGACTCTGG 503
  ||| ||| :::: ||::::: ||::::: ||::::: ||::::: ||:::::
185 CysLeuGlyPheLeuGlnGlyGlyLysAspSerCysGlnTyrAspSerGl 201
504 GGGGCCCTGATCTGCAACGGGTACTTGACAGGGCTGTGCTTTGGAA 553
  ||::::: ||::::: ||::::: ||::::: ||::::: ||::::: ||
201 yGlyProValValCysAsnGlyGluValGlnGlyIleValSerTrpGlyA 218
554 AAGCCCGCTGTGCCAACTGGCGTGGCAGGTGCTTACACCAACCTTGC 603
  :::: ||::: ||| ||::::: ||::::: ||::::: ||::::: ||
218 spGly...CysAlaLeuGlnGlyLysProGlyValTyrThrLysValCys 233
604 AAATTCACCTGAGTGATAGAGAAACCGTC 633
  ::::: ::::::::::: ||::::: ||::::: ||::::: ||::::: ||
234 AsnTyrLeuAsnTrpIleGlnGlnThrVal 243
seq_name: plr2:S33772
```

```
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) precursor, renal - crab-eating macaque
  C:Species: Macaca fascicularis (crab-eating macaque)
  C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 08-Sep-1997
  C:Accession: S33772
  R:Lin, F.K.; Lin, C.H.; Chou, C.C.; Chen, K.; Lu, H.S.; Bacheller, W.; Herrera, C.; J
  Biochim. Biophys. Acta 1173, 325-328, 1993
  A:Title: Molecular cloning and sequence analysis of the monkey and human tissue kall
  A:Reference number: S33772; MUID:93305727
  A:Accession: S33772
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-257 <LIN>
  A:Cross-references: GB:L10039; NID:g293140; PID:g293141
  A:Note: the authors translated the codon Cgr for residue 17 as Ala, GCG for residue 1
```


C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;25-249/Domain: trypsin homology <TRY>
F;62,116,209/Active site: His, Asp, Ser #status predicted

alignment_scores: Length: 223
 Quality: 461.50
 Ratio: 2.977 Gaps: 3
Percent Similarity: 69.507 Percent Identity: 40.359

alignment_block:
US-09-030-606-175 x S33772 ..

Align seg 1/1 to: S33772 from: 1 to: 257

```
1  GCGAGCCCTGGCAGGCGGCGACTGTCATGAAACGAATTGTTCTGCTC 50
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
32  SerGlnProTyrGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysG1 48
51  GGGGCTCCTGTGTCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCC 100
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
48  YGlyLeuValHisProGlnTyrValLeuThrAlaAlaHisCysTles 65
101 AGAAGCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAA 150
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
65  erAspAsnTyrGlnLeuTyrLeuArgHisAsnLeu...PheAspAsp 80
151 GAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTAAGGACCCAGA 200
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
81  GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProG1 97
201 GTACAACAGACTCTTGCTCGCTAAC..... 225
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
97  yPheAsnMetSerLeuLeuYsAsnHisThrArgGlnAlaAspAspTyrS 114
226 .....GACCTCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTCTGAC 270
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
114  erHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGluLeuThrAsp 130
271 ACCATCCGGAGCATCAGCATGCTTGCAGTGCCTACCGCGGGAGACTC 320
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
131  AlavalGlnValValGluLeuProThrGlnGluProGluValGlySerTh 147
321 TTGCTCTGTTCTGCTGGGTCTGCTG.....GCCAAGCGCAGAATGC 364
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
147  rCysLeuAlaSerGlyTyrGlySerIleGluProGluAsnPheSerPheP 164
365 CTACCGTGTGCTGCACTGCGTGAAGTGTGGTGTGTGAGGANGTCTGC 414
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
164  roAspAspLeuGlnCysValAspLeuGluLeuProAsnAspGluCys 180
415 AGTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCGCGG 464
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
181  AlalySaIaHisThrGlnLysValThrGluPheMetLeuCysAlaGlyH1 197
465 AGGCAAGACAGCAGAGACTCTGCAACGAGTGTGAGTGTGGGGGGCCCTGA 514
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
197  sleuGluGlyGlyLysAspThrCysValGlyAspSerGlyGlyProLeuT 214
515 TCTGCAACGGGCTACTTGCAGGCGCTTGTGCTTTCGAAAGCCCGCTGT 564
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
214  hrCysAspGlyValLeuGlnGlyValThrSerTyrGlyTyrIleProCys 230
565 GGGCAACTTGCGCTGCGAGGTGTCTACACCAACCTCTGCAATTCACATGA 614
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
231  GlySerProAsnLysProAlaValPheValLysValLeuSerTyrValLy 247
615 GTGATAGAGAAACCGTC 633
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
247  strPleGluAspThrIle 253
```

seq_name: p1r2:A31136

seq_documentation_block:

tissue kallikrein (EC 3.4.21.35) 7 precursor, submandibular - rat
N;Alternate names: glandular prokallikrein 7, submandibular; proteinase A
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 20-Mar-1998
C;Accession: A31136; S10698; S10699; D41429; B41429; S09315
R;Chen, Y.P.; Chao, J.; Chao, L.
Biochemistry 27, 7189-7196, 1988
A;Title: Molecular cloning and characterization of two rat renal kallikrein genes.
A;Reference number: A31136; MUID:89088074
A;Accession: A31136
A;Molecule type: DNA
A;Residues: 1-261 <CHE>
A;Cross-references: GB:M19647; GB:J02837; MID:g204999; PID:g205000
R;Elmoujaded, A.; Gutman, N.; Brillard, M.; Gauthier, F.
FEBS Lett. 265, 137-140, 1990
A;Title: Substrate specificity of two kallikrein family gene products isolated from t
A;Reference number: S10698; MUID:90306305
A;Accession: S10698
A;Molecule type: protein
A;Residues: 112-139 <EL2>
R;Kato, H.; Nakanishi, E.; Enjoji, K.; Hayashi, I.; Oh-Ishi, S.; Iwanaga, S.
J. Biochem. 102, 1389-1404, 1987
A;Title: Characterization of serine proteinases isolated from rat submaxillary gland:
A;Reference number: A41429; MUID:88198057
A;Accession: D41429
A;Molecule type: protein
A;Residues: 112-133 <KAT>
A;Accession: B41429
A;Molecule type: protein
A;Residues: 25-34, 'D', 36-45, 'S', 47-67, 'X', 69-75 <KA2>
R;Brady, J.M.; MacDonald, R.J.
Arch. Biochem. Biophys. 278, 342-349, 1990
A;Title: The expression of two kallikrein gene family members in the rat kidney.
A;Reference number: S09315; MUID:90225801
A;Accession: S09315
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 43-45, 'S', 47-114, 'A', 116-261 <BRA>
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-18/Domain: signal sequence #status predicted <Sig>
F;19-261/Product: tissue prokallikrein 7, submandibular #status predicted <MAT>
F;65,120,213/Active site: His, Asp, Ser #status predicted

alignment_scores: Length: 225
 Quality: 461.00
 Ratio: 2.918 Gaps: 3
Percent Similarity: 70.222 Percent Identity: 39.556

alignment_block:
US-09-030-606-175 x A31136 ..

Align seg 1/1 to: A31136 from: 1 to: 261

```
1  GCGAGCCCTGGCAGGCGGCGACTGTCATGAAACGAATTGTTCTGCTC 50
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
35  SerGlnProTyrGlnValAlaLeuTyrSerPheThrLysTyrLeuCysG1 51
51  GGGGCTCCTGTGTCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCC 100
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
51  YGlyValLeuIleAspProSerTyrPvalIleThrAlaAlaHisCysSers 68
101 AGAAGCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAA 150
   :::::::::::::::::::::  ::  ::  ::  ::  ::  ::  ::  ::
68  erAsnAsnTyrGlnValTyrPleuGlyArgAsnAsnLeuLeuGluAsp... 83
```


OM of: US-09-030-606-175 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL-frame+n2p.model -DEV-xlp
-O-/cgn2.1/USPTO_spool/US09030606/runat_24091999_171618_29883/app_query.fasta.1
-DB-SwissProt_37 -QFMT-fastan -SUFFIX-rsp -GAPOP-12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE-LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-175
Query length: 1167
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

score_list:

| Sequence | Strd Orig | ZScore | EScore | len | Documentation |
|---|-----------|---------|------------------------------|-----|---------------|
| SwissProt_37:SCCE_HUMAN + 530.50 551.14 1.5e-23 | 253 | P49862 | homo sapiens (human). | | |
| SwissProt_37:KLK1_RAT + 478.00 497.87 1.5e-20 | 244 | P36375 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_MOUSE + 472.00 491.28 3.2e-20 | 261 | P00756 | mus musculus (mouse) | | |
| SwissProt_37:TRY1_CHICK + 470.50 490.11 3.9e-20 | 248 | P090627 | gallus gallus (chicken) | | |
| SwissProt_37:SS9_HUMAN + 470.50 490.23 3.9e-20 | 244 | P092876 | homo sapiens (human) | | |
| SwissProt_37:TRY3_RAT + 469.50 489.12 4.4e-20 | 247 | P08426 | rattus norvegicus (rat) | | |
| SwissProt_37:TRY2_CHICK + 468.50 488.07 5.1e-20 | 248 | P090628 | gallus gallus (chicken) | | |
| SwissProt_37:TRY2_XENLA + 468.50 488.19 5.1e-20 | 244 | P07059 | xenopus laevis (afrod) | | |
| SwissProt_37:TRY3_CHICK + 467.50 487.05 5.8e-20 | 248 | P090629 | gallus gallus (chicken) | | |
| SwissProt_37:KLK8_RAT + 463.00 482.10 1.0e-19 | 261 | P36374 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK8_RAT + 463.00 482.16 1.0e-19 | 259 | P36376 | rattus norvegicus (rat) | | |
| SwissProt_37:TRYB_RAT + 462.00 481.50 1.2e-19 | 246 | P32822 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_MACFA + 461.50 480.68 1.3e-19 | 257 | P07276 | macaca fascicularis (macaca) | | |
| SwissProt_37:KLK7_RAT + 461.00 480.06 1.3e-19 | 261 | P36373 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_MOUSE + 461.00 480.06 1.3e-19 | 261 | P15946 | mus musculus (mouse) | | |
| SwissProt_37:TRY1_CANFA + 459.50 478.95 1.6e-19 | 246 | P06871 | canis familiaris (dog) | | |
| SwissProt_37:TRYA_RAT + 459.00 478.44 1.8e-19 | 246 | P32821 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_MOUSE + 458.00 477.00 2.0e-19 | 261 | P15948 | mus musculus (mouse) | | |
| SwissProt_37:KLK1_MOUSE + 458.00 477.06 2.0e-19 | 259 | P15948 | mus musculus (mouse) | | |
| SwissProt_37:TRY4_RAT + 457.00 476.38 2.3e-19 | 247 | P12788 | rattus norvegicus (rat) | | |
| SwissProt_37:TRY1_BOVIN + 456.50 475.98 2.4e-19 | 243 | P00760 | bos taurus (bovine) | | |
| SwissProt_37:KLK1_PAPHA + 456.00 475.05 2.6e-19 | 258 | P028773 | papio hamadryas (hama) | | |
| SwissProt_37:KLK9_MOUSE + 452.00 470.89 4.4e-19 | 261 | P15949 | mus musculus (mouse) | | |
| SwissProt_37:KLK2_MOUSE + 450.00 468.85 5.7e-19 | 261 | P36369 | mus musculus (mouse) | | |
| SwissProt_37:KLK2_MOUSE + 449.00 467.83 6.4e-19 | 261 | P20151 | homo sapiens (human) | | |
| SwissProt_37:PROS_HUMAN + 449.00 467.83 6.4e-19 | 261 | P07288 | homo sapiens (human) | | |
| SwissProt_37:PROS_HUMAN + 448.50 467.74 6.9e-19 | 246 | P07146 | mus musculus (mouse) | | |
| SwissProt_37:KLK2_RAT + 448.00 466.86 7.4e-19 | 259 | P00759 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_RAT + 447.00 465.79 8.4e-19 | 261 | P00758 | rattus norvegicus (rat) | | |
| SwissProt_37:KLK1_MOUSE + 446.00 464.77 9.5e-19 | 261 | P15947 | mus musculus (mouse) | | |
| SwissProt_37:TRY1_XENLA + 445.00 464.26 1.1e-18 | 243 | P19799 | xenopus laevis (afrod) | | |
| SwissProt_37:TRY1_RAT + 442.50 461.62 1.5e-18 | 246 | P00762 | rattus norvegicus (rat) | | |
| SwissProt_37:TRY3_SALSA + 442.50 461.86 1.5e-18 | 238 | P35033 | salmo salar (atlantic) | | |
| SwissProt_37:KLK5_MOUSE + 442.00 460.69 1.6e-18 | 261 | P15945 | mus musculus (mouse) | | |
| SwissProt_37:TRY2_BOVIN + 441.50 460.57 1.7e-18 | 247 | P029463 | bos taurus (bovine) | | |
| SwissProt_37:TRY2_PIG + 440.50 460.03 2.0e-18 | 231 | P00761 | sus scrofa (pig) | | |
| SwissProt_37:TRY2_CANFA + 438.50 457.52 2.6e-18 | 247 | P06872 | canis familiaris (dog) | | |
| SwissProt_37:TRY2_SQUAC + 438.00 457.54 2.7e-18 | 229 | P00764 | squalus acanthias (sp) | | |
| SwissProt_37:KLK6_MOUSE + 437.00 455.60 3.1e-18 | 261 | P00765 | mus musculus (mouse) | | |
| SwissProt_37:KLK1_HUMAN + 436.50 455.06 3.3e-18 | 262 | P06870 | homo sapiens (human) | | |
| SwissProt_37:PROS_MACMU + 435.00 453.56 4.0e-18 | 261 | P33619 | macaca mulatta (rhesu) | | |
| SwissProt_37:TRY1_HUMAN + 434.50 453.44 4.3e-18 | 247 | P07477 | homo sapiens (human) | | |

SwissProt_37:TRY2_SALSA + 434.00 453.40 4.6e-18 231 P35033 salmo salar (atlan
SwissProt_37:KLK9_RAT + 429.00 447.49 8.8e-18 259 P07647 rattus norvegicus (r
SwissProt_37:KLK2_CAVPO + 429.00 448.06 8.9e-18 239 P12323 cavia porcellus (g

seq_name: SwissProt_37:SCCE_HUMAN

seq_documentation_block:

| ID | SCCE_HUMAN | STANDARD; | PRT; | 253 AA. |
|----|--|-----------|------|---------|
| AC | P49862; | | | |
| DT | 01-OCT-1996 (REL. 34, CREATED) | | | |
| DT | 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE) | | | |
| DT | 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE) | | | |
| DE | STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE). | | | |
| GN | PRSS6 OR SCCE. | | | |
| OS | HOMO SAPIENS (HUMAN). | | | |
| OC | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; | | | |
| OC | PRIMATES; CATARRHINI; HOMINIDAE; HOMO. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 23-53. | | | |
| RC | TISSUE-SKIN; | | | |
| RX | MEDLINE; 94308225. | | | |
| RA | HANSSON L., STROMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A., | | | |
| RA | EGLERUD T.; | | | |
| RT | "Cloning, expression, and characterization of stratum corneum | | | |
| RT | chymolytic enzyme. A skin-specific human serine proteinase."; | | | |
| RL | J. BIOL. CHEM. 269:19420-19426(1994). | | | |
| RN | [2] | | | |
| RP | CHARACTERIZATION. | | | |
| RX | MEDLINE; 95314630. | | | |
| RA | SKYTT A., STROMQVIST M., EGLERUD T.; | | | |
| RT | "Primary substrate specificity of recombinant human stratum corneum | | | |
| RT | chymolytic enzyme."; | | | |
| CC | BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995). | | | |
| CC | -1- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE | | | |
| CC | STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS | | | |
| CC | SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID | | | |
| CC | RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE | | | |
| CC | CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE- | | | |
| CC | -1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION | | | |
| CC | OF PRECURSORS TO INFLAMMATORY CYTOKINES. | | | |
| CC | -1- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS | | | |
| CC | EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE | | | |
| CC | ALSO SEEN IN THE BRAIN AND KIDNEY. | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | | |
| CC | TRYPSIN FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; L33404; G532504; - | | | |
| DR | PROSITE; PS00134; TRYPSIN_HIS; 1. | | | |
| DR | PROSITE; PS00135; TRYPSIN_SER; 1. | | | |
| DR | PFAM; PF00089; trypsin; 1. | | | |
| DR | HSSP; P00763; IDPO. | | | |
| KW | HYDROLASE; SERINE PROTEASE; ZMOGEN; GLYCOPROTEIN; SIGNAL. | | | |
| FT | SIGNAL 1 22 | | | |
| FT | PROPEP 23 29 | | | |
| FT | CHAIN 30 253 | | | |
| FT | ACT_SITE 70 70 | | | |
| FT | ACT_SITE 112 112 | | | |
| FT | ACT_SITE 205 205 | | | |
| FT | ACT_SITE 207 207 | | | |
| FT | DISULFID 36 137 | | | |
| FT | DISULFID 55 71 | | | |
| FT | DISULFID 144 211 | | | |
| FT | DISULFID 176 190 | | | |
| FT | DISULFID 201 226 | | | |
| FT | CARBOHYD 246 246 | | | |
| SO | SEQUENCE 253 AA; 27525 MW; 07FDB9F7 CRC32; | | | |

alignment_scores:

Quality: 530.50 Length: 214
Ratio: 3.255 Gaps: 3
Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:

US-09-030-606-175 x SCCE_HUMAN ..

Align seg 1/1 to: SCCE_HUMAN from: 1 to: 253

```
1 GCGCAGCCCTGGCAGGCGGCGCAGTGCATGGAACGAATTGTTCTC 50
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|
40 SerHisProTprGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCys 56
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|
51 GGGCGTCTGGTGCATCCGCGAGTGGTGTCTGCAGCCGACACTGTTCC 100
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|
56 YGlyValLeuValAsnGluArgTprValLeuThrAlaAlaHisCysLys 73
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|
101 AGAAGCTCTACACCATCGGCTGGGCTGCACAGTCTTGAGCCGACCAA 150
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
73 etAsnGlnTyrThrValHisLeuGlySerAspThrLeu..GlyAspArg 88
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
151 GAGCCAGGAGCCAGATGTTGAGGCGCAGCTCTCCGTACGCGACCCAGA 200
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProGl 103
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
201 GTACACAGACTCTGCTCGCTACGACCTCATGCTCATCAAGTTGACG 250
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
103 YTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsn 120
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
251 AATCCGTGTCGAGTCTGACACACCATCCGAGCATCAGCATGCTTCGAC 300
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
301 TGCCCTACCGGGGGAAGTCTTGCTCGTCTGCTGGCTGGCTGCTGGC 350
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
137 CysGluProProGlyThrThrCysThrValSerGlyTprGlyThrThr 153
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
351 GAACGCGC.....AGATGCTACCGCTGCTGCACAGTGCAGCGTGC 394
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLys 170
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
395 TGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGACCACCC 444
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
170 euLeSerProGlnAspCysThrLysValTyrLysAspLeuGluAsn 186
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
445 AGCATGTTCTGCGCCCGCGGCGGAGGCGCAAGCAGAGGACTCTGCAAC 494
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
187 SerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsn 203
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
495 TGAAGTCTGGGGGGCCCGCTGATCTGCAAGGGGTACTTGAGGCGCTGT 544
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
203 YAspSerGlyCysProLeuValCysArgGlyThrLeuGlnGlyLeuVal 220
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
545 CTTTCGAAAAGCCCGCTGTGGCCCACTGGCGCTGCAGGTCTTACACC 594
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
220 erTprGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
595 AACCTCTGCAATTCAGTGAAGTAGAGAAACCGCTCCAG 636
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
237 GlnValCysLysPheThrLysTprLeuAsnAspThrMetLys 250
```

seq_name: SwissProt_37:KLKA_RAT

seq_documentation_block:

ID KLKA_RAT STANDARD; PRT; 244 AA.
AC P36375;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 10 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)

DE (T-KININOGENASE) (K10) (PROTEINASE B) (ENDOPEPTIDASE K) (FRAGMENT).
GN KLK10 OR KLK-10.
OS RATUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-SUBMANDIBULAR GLAND, AND KIDNEY;
RX MEDLINE; 93041794.
RA MA J.-X., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of rklk10, a cDNA encoding T-
kininogenase from rat submandibular gland and kidney.";
RL BIOCHEMISTRY 31:10922-10928(1992).
RN [2]
RP SEQUENCE OF 10-32; 95-124 AND 179-232.
RX MEDLINE; 91224135.
RA GUTMAN N., ELMOUTAHED A., BRILLARD M., DU SORBIER B., GAUTHIER F.;
RT "Microheterogeneity of rat submaxillary gland kallikrein k10, a
member of the kallikrein family.";
RL EUR. J. BIOCHEM. 197:425-429(1991).
RN [3]
RP SEQUENCE OF 10-32 AND 97-133.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE; 90153911.
RA XIONG W., CHEN L.-M., CHAO J.;
RT "Purification and characterization of a kallikrein-like
T-kininogenase.";
RT J. BIOL. CHEM. 265:2822-2827(1990).
RN [4]
RP SEQUENCE OF 10-32 AND 97-117.
RX MEDLINE; 88198057.
RA KATO H., NAKANISHI E., ENYIOJI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT "Characterization of serine proteinases isolated from rat
submaxillary gland: with special reference to the degradation of rat
kininogens by these enzymes.";
RL J. BIOCHEM. 102:1389-1404(1987).
CC -I- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS PROTEIN MAY BE
INVOLVED IN THE REGULATION OF RENAL FUNCTION.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-I-XAA BONDS IN
SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
MET-I-XAA OR LEU-I-XAA.
CC -I- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED
BY A DISULFIDE BOND.
CC -I- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND, WHERE IT IS
FOUND IN THE GRANULAR CONVOLUTED TUBULE AND STRIATED DUCT CELLS.
CC IT IS LIKELY THAT THE ENZYME IS MAINLY SYNTHESIZED IN THE GRANULAR
CONVOLUTED TUBULES AND THEN TRANSFERRED TO OTHER TISSUES BY
RELEASE INTO THE VASCULATURE OR INTERSTITIAL SPACE.
CC -I- PTM: PROBABLY N- AND O-GLYCOSYLATED. IT HAS CARBOHYDRATE MOIETIES
OF ALPHA-METHYL-D-MANNOSIDE AND N-ACETYL-D-GLUCOSAMINE GROUPS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; S48142; G259431; .
DR PIR; A35545; A35545.
DR PIR; B35545; B35545.
DR PIR; A44284; A44284.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00759; lton.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; SIGNAL;
KW MULTIGENE FAMILY.


```

FT  NON_TER      1      1      PROBABLE.
FT  SIGNAL      <1      3
FT  PROPEP      4      9      ACTIVATION PEPTIDE (PROBABLE).
FT  CHAIN      10      244      GLANDULAR KALLIKREIN 10.
FT  CHAIN      10      96      T-KININOGENASE LIGHT CHAIN.
FT  CHAIN      97      244      T-KININOGENASE HEAVY CHAIN.
FT  ACT_SITE     48      48      CHARGE RELAY SYSTEM.
FT  ACT_SITE     103     103      CHARGE RELAY SYSTEM.
FT  ACT_SITE     196     196      CHARGE RELAY SYSTEM.
FT  DISULFID     16      156      BY SIMILARITY.
FT  DISULFID     33      49      BY SIMILARITY.
FT  DISULFID     135     202      BY SIMILARITY.
FT  DISULFID     167     181      BY SIMILARITY.
FT  DISULFID     192     217      BY SIMILARITY.
FT  CARBOHYD     76      76      POTENTIAL.
FT  CARBOHYD     91      91      POTENTIAL.
FT  CARBOHYD     126     126      POTENTIAL.
FT  CARBOHYD     142     142      POTENTIAL.
FT  CONFLICT     29      29      N -> IET (IN REF. 3).
FT  CONFLICT     115     116      IT -> DS (IN REF. 4).
FT  CONFLICT     128     128      E -> G (IN REF. 3).
FT  CONFLICT     133     133      S -> G (IN REF. 3).
SQ  SEQUENCE     244 AA; 27305 MM; 7661F5F CRC32;
```

```

alignment_scores:
  Quality: 478.00      Length: 225
  Ratio: 2.987      Gaps: 4
  Percent Similarity: 71.111      Percent Identity: 41.333
```

alignment_block:
US-09-030-606-175 x KLKA_RAT ..

Align seg 1/1 to: KLKA_RAT from: 1 to: 244

```

1  GCGCAGCCCTGGCAGCGGCACACTGGTCATGGAAGAATGTTCTGCTC 50
   ::::::::::::::::::::: ||||| :::::::
20 SerGlnProTrrpGlnValAlaIleIle.....AsnGluTyrLeucysgl 34
51 GGGCGTCTGTGTCATCCGCAGTGGGTGCTGTCAGCCGACACTGTTCC 100
   ::::::::::::::: ||::::: ||::::: ||:::::
34 yGlyValLeuIleAspProSerTrrpValIleThrAlaIaHisCystYrs 51
101 AGAAGCTCTACACCAATCGGCTGGGCTGCAGTCTTGAGGCCGACCAA 150
   ::||| ||| :::: ||||| ::::::||| |||
51 erAsnTyrTrrHisValLeuLeuGlyArgAsnAsnLeuPheGluasp... 66
151 GAGCCAGGAGCCAGATGGTGAGGCCAGCTCTCCGTACGGCAGCCAGA 200
   ||||| ::||| :::::: ||| ||||| ::
67 GluProPheAlaGlnTyrArgPheValAsnGlnSerPheProHisProAs 83
201 GTACAACAGACTCTGCTC..... 219
   ::||| ::||| ::||| ::||| ::||| ::|||
83 pTyrLysProPheLeuMetArgAsnHisThrArgGlnArgGlyAspAspT 100
220 ..GCTACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCT 267
   ::||| ||||| ||||| ::||| ::||| ::||| ::|||
100 ySerAsnAspLeuMetLeuLeuHisLeuSerGluProAlaAspIleThr 116
268 GACACCATCCGAGACATCAGCATTTGCTGCGAGTGCCTACCGCGGGA 317
   ||| :::::: ||::::: ||| ::||| ::
117 AspGlyValValLeuValLeuAspLeuProThrGluGluProLysValGlySe 133
318 CTCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
   ::||| ||||| ::||| ||||| ||||| |||||
133 rThrCysLeuAlaSerGlyTrrpGlySerThrLysProLeuAsnTrrpGluL 150
362 TGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
   ::||| ||||| ||||| ||||| ::||| ::||| ::|||
150 euProAspAspLeuGlnCysValAsnIleHisLeuLeuSerAsnGluLys 166
412 TGCAGTAGCTCTATGACCCGCTGTACACCCAGCATGTCTGCGCGG 461
   ||| :::: ||||| ||||| ||||| |||||
```

```

167 CysIleGluAlaTyrGlnGlnLysValThrAspLeuMetLeuCysAlaG1 183
462 CGGAGGCAAGACAGACAGAGACTCTGCAACGGTGTACTCTGGGGGCCCC 511
   :::: ::||| ::||| ::||| ::||| ::|||
183 yGluMetAspGlyArgLysAspThrCysLysGlyAspSerGlyGlyProL 200
512 TGATCTGCACGGGTACTTGACAGGGCCTTGCTGCTTTCGGAAGACCCCG 561
   ||||| ::||| ||||| ::||| ::||| ::|||
200 euIleCysAspGlyValLeuGlnGlyIleThrSerTrrpGlyAsnValPro 216
562 TGTGCCCACTTGCGCTGCCAGGTGTCTACACCAACCTCTGCAATTCAC 611
   ||::::: ||||| ||||| ||||| |||||
217 CysAlaGluProTyrAsnProGlyValTyrThrLysLeuIleLysPheTh 233
612 TGAGTGATAGAGAAACCGTCCAG 636
   |::::| ::||| ::||| ::||| ::|||
233 rSerTrrpIleLysGluValMetLys 241

seq_name: SwissProt_37:KLK3_MOUSE
seq_documentation_block:
ID  KLK3_MOUSE      STANDARD;      PRT;      261 AA.
AC  P00756;
DT  21-JUL-1986 (REL. 01, CREATED)
DT  21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  GLANDULAR KALLIKREIN K3 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE  (MGR-3) (7S NERVE GROWTH FACTOR GAMMA CHAIN) (GAMMA-NGF).
GN  KLK3 OR KLK-3 OR NGFG.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 85076169.
RA  ULBRICH A., GRAY A., WOOD W.I., HAYFLICK J., SEEBURG P.H.;
RT  "Isolation of a cDNA clone coding for the gamma-subunit of mouse
RT  nerve growth factor using a high-stringency selection procedure.";
RL  DNA 3:387-392(1984).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 85257431.
RA  EVANS B.A., RICHARDS R.I.;
RT  "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT  are contiguous.";
RL  EMBO J. 4:133-138(1985).
RN  [3]
RP  SEQUENCE OF 25-261.
RX  MEDLINE; 81264363.
RA  THOMAS K.A., BAGLAN N.C., BRADSHAW R.A.;
RT  "The amino acid sequence of the gamma-subunit of mouse submaxillary
RT  gland 7 S nerve growth factor.";
RL  J. BIOL. CHEM. 256:9156-9166(1981).
RN  [4]
RP  X-RAY CRYSTALLOGRAPHY (3.15 ANGSTROMS) OF 7S COMPLEX.
RC  STRAIN-SWISS WEBSTER; TISSUE-SUBMAXILLARY GLAND;
RX  MEDLINE; 98035451.
RA  BAX B., BLUNDELL T.L., MURRAY-RUST J., McDONALD N.Q.;
RT  "Structure of mouse 7S NGF: a complex of nerve growth factor with
RT  four binding proteins.";
RL  STRUCTURE 5:1275-1285(1997).
CC  -1- FUNCTION: 7S NGF ALPHA CHAIN STABILIZES THE 7S COMPLEX. THE BETA
CC  DIMER PROMOTES NEURITE GROWTH. THE GAMMA CHAIN IS AN ARGININE-
CC  SPECIFIC PROTEASE; IT MAY ALSO HAVE PLASMINOGEN ACTIVATOR
CC  ACTIVITY, AS WELL AS MITOGENIC ACTIVITY FOR CHICK EMBRYO
CC  FIBROBLASTS.
CC  -1- SUBUNIT: 7S NERVE GROWTH FACTOR IS COMPOSED OF TWO ALPHA CHAINS,
CC  A BETA DIMER COMPOSED OF IDENTICAL CHAINS, AND TWO GAMMA CHAINS.
CC  -1- THIS PRECURSOR IS CLEAVED INTO SEGMENTS TO PRODUCE THE ACTIVE
CC  FORM OF THE GAMMA CHAIN, WHICH OCCURS NATURALLY AS COMBINATIONS
CC  OF EITHER TWO OR THREE SEGMENTS HELD TOGETHER BY DISULFIDE
CC  BONDS: B1 + A OR B1 + C + B2.
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
```



```
CC -----
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01389; G53374; -
DR EMBL; X01798; G582429; -
DR EMBL; X01799; E73575; -
DR PIR; A00942; NMSG.
DR PDB; 1SGF; 27-MAY-98.
DR MGI; 97322; NGFG.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PFAM; PF00089; trypsin; 1.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZMOGEN; SIGNAL; GROWTH FACTOR; 3D-STRUCTURE.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE.
FT CHAIN 25 261 GLANDULAR KALLIKREIN K3.
FT CHAIN 25 107 NERVE GROWTH FACTOR GAMMA CHAIN 1.
FT CHAIN 112 261 NERVE GROWTH FACTOR GAMMA CHAIN 2.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173
FT DISULFID 50 66
FT DISULFID 152 219
FT DISULFID 184 198
FT DISULFID 209 234
FT CARBOHYD 102 102
FT DOMAIN 25 107 SEGMENT B1.
FT DOMAIN 112 261 SEGMENT A.
FT DOMAIN 112 164 SEGMENT C.
FT DOMAIN 165 261 SEGMENT B2.
FT CONFLICT 108 111 MISSING (IN REF. 2).
SQ SEQUENCE 261 AA; 28998 MW; 7B9C5256 CRC32;
```

```
alignment_scores:
  Quality: 472.00      Length: 228
  Ratio: 2.969        Gaps: 4
Percent Similarity: 69.737 Percent Identity: 39.912
```

alignment_block:

us-09-030-606-175 x KLK3_MOUSE ..

Align seg 1/1 to: KLK3_MOUSE from: 1 to: 261

```
1 GCGCAGCCCTGGCAGGCGGCGACATGGAAGAAGATGTCTGCTC 50
  :::::::::::::::::::::
35 SerGlnProItrPhisValAlaValAlaValTyrTyrThrGlnTyrLeuGlySgl 51
51 GGGCGTCTGGTGATCCGCGAGTGGGTGCTGCAGCCGCGACACTGTTCC 100
  :::::::::::::::::::::
51 yGlyValLeuLeuAspProAsnTyrValLeuThrAlaAlaHisCysTyrA 68
101 AGAAGCTCTAGACCATGGGCTGGGCTGCACAGTCTTGAGGCCGACCAA 150
  :::::::::::::::::::::
68 spAspAsnTyrLysValTyrPleuGlyLysAsnAsnLeuPheLysAsp... 83
151 GAGCCAGGAGCCAGATGTGAGAGCCAGCCTCTCCGTAGCCAGCCAGCA 200
  |||||:|||||
84 GluProSerAlaGlnHisArgPheValSerLysAlaIleProHisProG1 100
201 GTACAACAGACTCTGTGTC..... 219
  ::|||
100 yPheAsnMetSerLeuMetArgLysHisIleArgPheLeuGluTyrAspT 117
```

```
220 ..GCTAAGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCGAGTCT 267
  :::::::::::::::::::::
117 ySerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIleThr 133
268 GACACCATCCGGAGCATCATGCTGTCGAGTGGCCCTACCGGGGGA 317
  |||||:|||||
134 AspIleValLysProIleThrLeuProThrGluGluProLysLeuGlySe 150
318 CTCTGCTCTGCTGCTGCTGGGTCTGCTGCGCAACGGCAGATGCTTA 367
  :::::::::::::::::::::
150 rThrcysLeuAlaSerGlyTyrGlySerIleThr.....Prot 163
368 CCGTG.....CTGCACCTGCGTGAACGTGTCGGTGTG 399
  ||
163 hrlYsPheGlnPheThrAspAspLeuTyrCysValAsnLeuLysLeu 179
400 TCTGAGGANGTCTGAGTAAGCTCTATGACCCGCTGTACACCCAGCAT 449
  ::
180 ProAsnGluAspCysAlaLysAlaHisIleGluLysValThrAspAlaMe 196
450 GTTCGCGCCGGCGGAGGCAAGACAGAGACTCCTGCAACGGTACT 499
  :::::::::::::::::::::
196 tLeuCysAlaGlyGluMetAspGlyGlyLysAspThrCysLysGlyAsp 213
500 CTGGGGGGCCCTGATCTGCAACGGGTACTTGACGGGCTGTGTCTTC 549
  |||||:|||||
213 ergLysGlyProLeuIleCysAspGlyValLeuGlnGlyIleThrSerTrp 229
550 GGAAGACCCCGTGTGGCCACTTGCGCTGCCAGTGTCTACACCACT 599
  |||
230 GlyHisThrProCysGlyGluProAspMetProGlyValTyrThrLysIle 246
600 CTGCAATTCAGTGAATGAGAAAACCGTC 633
  |
246 uAsnLysPheThrSerTrpIleLysAspThrMet 257
```

seq_name: SwissProt_37:TRY1_CHICK

seq_documentation_block:

ID TRY1_CHICK STANDARD; PRT; 248 AA.

AC Q90627;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRYPsinOGEN I-P1 PRECURSOR.

OS GALLUS GALLUS (CHICKEN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;

OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-PANCREAS;

RX MEDLINE; 95251611.

RA WANG K., GAN L., LEE I., HOOD L.E.;

RT "Isolation and characterization of the chicken trypsinogen gene

family.";

RL BIOCHEM. J. 307:471-479(1995).

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE

CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U15155; G603903; -

DR PROSITE; PS00134; TRYPsin_HIS; 1.


```

FT PROPEP 17 21 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 22 244 PROTEASE M.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 106 106 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 157 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 131 231 BY SIMILARITY.
FT DISULFID 138 203 BY SIMILARITY.
FT DISULFID 168 182 BY SIMILARITY.
FT DISULFID 193 218 BY SIMILARITY.
FT CARBOHYD 134 134 POTENTIAL.
SQ SEQUENCE 244 AA; 26856 MW; 0CFCECFE CRC32;
```

alignment_scores: Quality: 470.50 Length: 215
 Ratio: 3.055 Gaps: 3
Percent Similarity: 71.628 Percent Identity: 42.326

alignment_block:

US-09-030-606-175 x PSS9_HUMAN ..

Align seg 1/1 to: PSS9_HUMAN from: 1 to: 244

```

1 GCGCAGCCCTGGCAGCGCGCACTGTCATGAAACGAATTGTTCTGCTC 50
  :::::::::::::::::::::
32 SerHisProTyrGlnAlaIleuTyrThrSerGlyHisLeuLeuCysG 48
51 GGGCGTCTGTGTCATCCGCGAGTGGGTGCTGTACGCCGACACTGTTCC 100
  :::::::::::::::::::::
48 yGlyValLeuIleHisProLeuTyrValLeuThrAlaAlaHisCysLysL 65
101 AGAACTCCTACACCATCGGCGCTGGCCCTGCACAGTCTT.....GAG 141
  :::::::::::::::::::::
65 ySProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu 81
142 GCCGACCAAGAGCCAGGCGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACG 191
  :::::::::::::::::::::
82 SerSerGlnGluGln...SerSerValValArgAla.....ValIle 94
192 GCACCCAGAGTACACAGACTCTTGCTGCTACGACCTCATGCTCATCA 241
  :::::::::::::::::::::
94 eHisProAspTyrAspAlaAlaSerHisAspGlnAspIleMetLeuLeuA 111
242 AGTTGACGAATCCGTGCTCCGAGTCTGTGACACACCACCGGAGCATCAGCAT 291
  :::::::::::::::::::::
111 rGleuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127
292 GCTTCGAGTGCCTACCGCGGGGAACTCTTGCCCTGCTGCTGCTGGGG 341
  :::::::::::::::::::::
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTyrG 144
342 TCTGCTGGCGAAGCGGAGATGCCCTACCGTGTGCTGCTGCTGCTGCTGCTG 391
  :::::::::::::::::::::
144 yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTyrIleH 161
392 CGGTGGTGTCTGAGGANGTCTGCAAGTAAAGCTATGACCCGCTGTACCAC 441
  :::::::::::::::::::::
161 lSerValSerArgGluGluCysGluHisAlaTyrProGlyGlnIleThr 177
442 CCCAGCATGTTCTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 491
  :::::::::::::::::::::
178 GlnAsnMetLeuCysAlaGlyAspGlyLysTyrGlyLysAspSerCysG 194
492 CGGTGACTCTGGGGCGCGCTGATCTGCAACGGGCTGCTGAGGGCCTTG 541
  :::::::::::::::::::::
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeuV 211
542 TGTCTTTGGAAGCCCGCTGTGGCCAACCTGGCGTGGCCAGGTGTCTAC 591
  :::::::::::::::::::::
211 aISerTyrGlyAsnIleProCysGlySerLysGluLysProGlyValTyr 227
```

```

592 ACCAACCTCTGCAATTCACGTAGATAGAGAAACCGTCCAG 636
  :::::::::::::::::::::
228 ThrAsnValCysArgTyrThrAsnTyrIleGlnLysThrIleGln 242
```

seq_name: SwissProt_37::TRY3_RAT

seq_documentation_block: ID TRY3_RAT STANDARD; PRT; 247 AA.

AC P08426;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN III, CATIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN DE III).
OS EUTRYS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATIUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87271609.
RA FLETCHER T.S., ALHADEFF M., CRAIK C.S., LARGMAN C.;
RT "Isolation and characterization of a cDNA encoding rat cationic trypsinogen."
RL BIOCHEMISTRY 26:3081-3086(1987).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; M16624; G206499; -.
DR PIR; A27547; A27547.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; 1DPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 15
FT PROPEP 16 24
FT CHAIN 25 247
FT ACT_SITE 64 64 TRYPSIN III, CATIONIC.
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 161 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 133 234 BY SIMILARITY.
FT DISULFID 140 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 197 221 BY SIMILARITY.
FT SITE 195 195 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 247 AA; 26269 MW; 7DD35F0B CRC32;
```

alignment_scores: Quality: 469.50 Length: 211
 Ratio: 2.953 Gaps: 4
Percent Similarity: 75.355 Percent Identity: 41.706

alignment_block:

US-09-030-606-175 x TRY3_RAT ..

Align seg 1/1 to: TRY3_RAT from: 1 to: 247

```

7 CCTGGCAGCGCGCACTGTCATGAAACGAATTGTTCTGCTGGCGCT 56
  :::::::::::::::::::::
  :::::::::::::::::::::
```



```

37 ProTyrGlnValSerLeu...AsnAlaGlyTyrHisPheCysGlyGlySe 52
157 CTGTGTCATCCGCAGTGGGTGCTGTACGCCGACACTTTCCAGA ACT 106
|||::: |||||||:::|||||:::|||||:::|||||:::
52 rleuileasnserGlnProValValSerAlaAlaHisCysTyrLysSera 69
85 GlyGlnGlnPheIleaspAlaAlaLysIleIleArgHisProSerTyrAs 101
107 CCTACACCATCGGGCTGGGCTTGACAGTCTTGAGGCCGACCAAGACCA 156
::: ||||| |||:::|||||::: |||
69 rGleGlnValArgLeuGlyGlnHisAsnIleaspVal...ValGlnGly 84
157 GGGAGCCAGATGGTGGAGGCCAGCCCTCTCCGTACGGCACCCAGAGTACA 206
|||:::|||||:::|||||::: ::||| |||||:::|||||
85 GlyGlnGlnPheIleaspAlaAlaLysIleIleArgHisProSerTyrAs 101
207 CAGACTCTTGCTCGCTAACAGACCTCATGCTCATCAAGTTGAGCAATCCG 256
|::: |||||:::|||||:::|||||:::|||||:::
101 nalaasnthrPheaspnaspilemetleuileLysleuasnserserProA 118
257 TGTCCGAGTCTGACACCAATCCGGAGCATCAGCATTGCTTCGAGTGCCT 306
::: |||||::: |||::: |||||::: |||
118 laThrleuasnserserArgValSerThrValSerLeuproArgSerCysGly 134
307 ACCCGGGGAACCTCTGCTCTGNTCTGGCTGGGT....CTGCTGGC 350
::::|||:::|||||:::|||||:::|||||::: |||
135 serserglythrLyscysleuvalserglytyrGlyasnThrleusense 151
351 GAACGGCAGAAATGCTACCGTGTCTGCACTGCGTGAACSTGTGCTGTGT 400
::: ||::: |||:::|||||:::|||||:::|||||::: |||:::|
151 rGlyThrASNtyrProserleuengInCysleuaspaIalProValleus 168
401 CTGAGGANGTCTGACATACTCTATGACCCGCTGTACCAACCCAGCATG 450
|||::: |||::: |||
168 eraSpserSerCysLysSerSerTyrProGlyLysIleThrSerasmet 184
451 TTCTGCGCGCGCGGAGGCGCAAGACCAAGAGACTCTGCAACGGTGACTC 500
||||| ||| ::| |||||||:::|||||
185 pheCysleuglYphelenglYglYglYlysaspserCysgInglyaspse 201
501 TGGGGGGCCCCGTGATCTGCAACGGGTACTTGACAGGCGCTGTGCTTTGCG 550
||||| |||::: |||||::: |||||::: |||||::: |||||:::
201 rGlyGlyProValValCysasnGlyInleugInglyValValSerTrpg 218
551 GAAAAAGCCCGTGTGGCCAATTGGCGTGCAGAGTGTCTACACCAACCTC 600
||::: |||::: ||| ||| ||||||| |||||::: |||
218 lYTyrGly..CysAlaGlnLysGlyLysProGlyValTyrThrLysVal 233
601 TGCAAATTCACCTAGTGCATAGAGAAAACCGTC 633
|||::: |||::: |||||::: |||||::: |||||::: |||||
234 CysasnTyrrValasnTrpIleGlnGlnThrval 244

seq_name: SwissProt_37:TRY2_CHICK

ID TRY2_CHICK STANDARD; PRT; 248 AA.
AC Q90628;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPsinogen I-P38 PRECURSOR.
OS GALUS Gallus (CHICKEN).
OC EUKARYOTA; METAzoA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RX MEDLINE; 95251611.
RA WANG K., GAN L., LEE I., HOOD L.E.;
RT "Isolation and characterization of the chicken trypsinogen gene family.";
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
```

```

CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U15156; G603905; -
CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
CC DR PFAM: PF00089; trypsin; 1.
CC DR HSSP: P00763; IDPO.
CC DR HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KM MULTIGENE FAMILY.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPSIN I-P38.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 248 AA; 26087 MW; F64E0643 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 468.50 | Length: 216 |
| Ratio: | 2.910 | Gaps: 5 |
| Percent Similarity: | 74.537 | Percent Identity: 43.056 |

alignment_block:

US-09-030-606-175 x TRY2_CHICK

Align seg 1/1 to: TRY2_CHICK from: 1 to: 248

```

1 GCGCAGCCCTGGCAGGCGGCACTGGTCATGAAACGAATTGTTCTGCTC 50
  ||| |||::|||:::||||| ::::: |||||:::
36 AlaAlaProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysG1 51
51 GGGCGTCTTGTCATCCGCAGTGGTCTGTCAGCCGCACACTGTTCC 100
  :||| |||::: ||||| ||||| ||||| ||||| ||||| ||||| :
51 yGlySerLeuIleSerSerGlnTrpValLeuSerAlaAlaHisCysTyrL 68
101 AGAAGCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCA 150
  ::::: ||| ::: ||||| ::::: ||| ||| |||
68 ySerSerIleGlnValLysLeuGlyGluTyrAsnLeuAlaAla...Gln 83
151 GAGCCAGGAGCCAGATGTTGGAGGCCAGCCCTCCGTACCGCACCCAGA 200
  :: ::::: ||| ::::: ||| ::||| |||
84 AspGlySerGluGlnThrIleSerSerLysValIleArgHisSerG1 100
201 GTACACAGAGACTTGTGCTCGCTAACGACCTCATGCTCATCAAGTTGACG 250
  ||||| ||| |||||::: ||||| ||||| ||||| :::::
100 yTyrAsnAlaAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuSerL 117
251 AATCCGTTCCGAGTCTGACACCAATCCGGAGCATCAGCATTTGCTTCGAG 300
  ::::: ||| ::::: ||||| ::: :::::
117 ySAlaIaIaThrLeuAsnSerTyrValAsnThrValProLeuProThrSer 133
301 TGCCCTACCGCGGGGAAGCTCTTGCTCGTNTCTGGCTGGGGT...CTGCT 347
  ||| ||||| ||||| ::::: ||||| ||||| ||||| ||||| |||||
134 CysValThrAlaGlyThrThrCysLeuIleSerGlyTyrIleGlyAsnThrLe 150

```



```

348 GCGCAACGGCAGATG...CTACCGTGTGCTGCTGCGTGAACGTGTCG 394
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
150 userSerGlySerLeuYrProAspValLeuGlnCysLeuAsnAlaProy 167
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
395 TGGTGTCTGAGGANGTCTGCAGTAAAGCTCTATGACCCGCTGACCACCC 444
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
167 alLeuSerSerSerGlnCysSerSerAlaTyrProGlyArgIleThrSer 183
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
445 AGCATGTCTTGCAGCCGCGGAGGCAAGACCAAGACTCTGCAACGG 494
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
184 AsnMetIleCysIleGlyTyrLeuAsnGlyGlyLysAspSerCysGlnG 200
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
495 TGACTCTGGGGGCCCCCTGATCTGCAACGGGTACTTGCAAGCCCTGTGT 544
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
200 yAspSerGlyGlyProValAlaCysAsnGlyGlnLeuGlnGlyPheValS 217
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
545 CTTTCGGAAGAACCCCGTGTGCGCAACTTGCGGTGCCAGGTGTCTACAC 594
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
217 erTrrpGly...IleGlyCysAlaGlnLysGlyTyrProGlyValTyrThr 232
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
595 AACCTCTGCAAAATTCAGTGAAGTAGAGAAACCGTCCAGNCCAGT 642
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
233 LysValCysAsnTyrValSerTrrpIleLysThrThrMetSerSerAsn 248
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

seq_name: SwissProt_37:TRY2_XENLA

seq_documentation_block:

```

ID TRY2_XENLA STANDARD; PRT; 244 AA.
AC P70059;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN PRECURSOR.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG R., LYTLE L., GAN L., HOOD L.E.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U72330; G1621633; -
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFAM: PF00089; trypsin; 1.
DR HSSP: P00763; IDPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 21 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 22 244 TRYPSIN.
FT ACT_SITE 61 61 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 105 105 CHARGE RELAY SYSTEM (BY SIMILARITY)..
FT ACT_SITE 198 198 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 158 BY SIMILARITY.
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 130 231 BY SIMILARITY.
FT DISULFID 137 204 BY SIMILARITY.
FT DISULFID 169 183 BY SIMILARITY.
FT DISULFID 194 218 BY SIMILARITY.
FT SITE 192 192 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

```

SEQ SEQUENCE 244 AA; 26079 MW; 754B2E0E CRC32;

alignment_scores:

| Quality: | 468.50 | Length: | 216 |
|---------------------|--------|-------------------|--------|
| Ratio: | 2.947 | Gaps: | 4 |
| Percent Similarity: | 73.611 | Percent Identity: | 42.130 |

alignment_block:

US-09-030-606-175 x TRY2_XENLA ..

Align seg 1/1 to: TRY2_XENLA from: 1 to: 244

```

1 GCGCAGCCCTGGCAGCGCGGCACTGTGTCATGGAAGAAACGAATTTGCTGCTC 50
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
32 AlaValProTyrGlnValSerLeu...AsnAlaGlyTyrHisPheCysG 47
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
51 GGGCGTCTGTGTCATCCGAGTGGGTGCTGTGACGGCAGACTGTTTC 100
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
47 yGlySerLeuIleAsnSerGlnTrrpValValSerAlaAlaHisCysTyrL 64
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
101 AGAAGCTCTACACATCGGGCTGGGCTGCACAGTCTTGAGCGGACCAA 150
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
64 ySererArgIleGlnValArgLeuGlyGlnHisAsnIle...AlaLeuAsn 79
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
151 GAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCAGCCAGA 200
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
80 GluGlyThrGlnGlnPheIleAspSerGlnLysValIleLysHisProAs 96
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
201 GTACACAGACTCTTGTCTGCTAAGCACTCATGCTCATCAAGTTGGACG 250
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
96 nTyrAsnSererArgAsnLeuAspAsnAspIleMetLeuIleLysLeuSerT 113
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
251 AATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCATGTGCTTGCAG 300
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
113 hThrAlaArgLeuSerAlaAsnIleGlnSerValProLeuProSerAla 129
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
301 TGCCCTACCGCGGGGCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
130 CysAlaSerAlaGlyThrAsnCysLeuIleSerGlyTrrpGlyAsnThrLe 146
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
345 GCTGCGAAGCGCAAGATGCTTACCGTGTGCACTGCGTGAACGTGTCGG 394
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
146 userSerGlyThrAsnTyrProAspLeuLeuGlnCysLeuAsnAlaProI 163
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
395 TGGTGTCTGAGGANGTCTGCAGTAAAGCTCTATGACCCGCTGACCACCC 444
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
163 leuThrAspSerGlnCysSerSerAsnSerTyrProGlyGlnIleThrLys 179
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
445 AGCATGTCTTGCAGCCGCGGAGGCAAGACCAAGAGACTCTGCAACGG 494
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
180 AsnMetPheCysAlaGlyPheLeuAlaGlyGlyLysAspSerCysGlnG 196
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
495 TGACTCTGGGGGCCCCCTGATCTGCAACGGGTACTTGCAAGGCCCTGTGT 544
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
196 yAspSerGlyGlyProValAlaCysAsnGlyGlnLeuGlnGlyValValS 213
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
545 CTTTCGGAAGAACCCCGTGTGCGCAACTTGCGGTGCCAGGTGTCTACAC 594
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
213 erTrrpGlyTyrGly...CysAlaGlnArgAsnTyrProGlyValTyrThr 228
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
595 AACCTCTGCAAAATTCAGTGAAGTAGAGAAACCGTCCAGNCCAGT 642
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
229 LysValCysAsnPhenValThrTrrpIleGlnSerThrIleSerSerAsn 244
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

seq_name: SwissProt_37:TRY3_CHICK

seq_documentation_block:

```

ID TRY3_CHICK STANDARD; PRT; 248 AA.
AC Q90629;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

```



```

DE TRYPsinogen II-P29 PRECURSOR.
OS GALUS GALLUS (CHICKEN).
OC EUKARYOTA; METAzoA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=pancreas;
RX MEDLINE; 95251611.
RA WANG K., GAN L., LEE I., HOOD L.E.;
RT "Isolation and characterization of the chicken trypsinogen gene
family.";
RL BIOCHEM. J. 307:471-479(1995).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U15157; G603907;
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; IDPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KM MULTIGENE FAMILY.
FT SIGNAL 1 16 BY SIMILARITY.
FT PROPEP 17 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPsin II-P29.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 248 AA; 26622 MW; F7FE97E9 CRC32;

alignment_scores:
      Quality: 467.50      Length: 211
      Ratio: 2.959      Gaps: 5
Percent Similarity: 74.882      Percent Identity: 42.654

alignment_block:
US-09-030-606-175 x TRV3_CHICK ..

Align seg 1/1 to: TRV3_CHICK from: 1 to: 248

      7 CCCTGGCAGCGGCGACTGTCATGGAAGCAATGTTCTGCTCGGCGGT 56
      |||::|||:::|||||      ::::: |||||:::|||||
      38 ProTyglnValSerLeu...AsnSerGlyTyRHisphecGlyGlySe 53
      57 CCTGTGCATCGCAGTGGGTGCTGTACGCCGACACACTGTTCCAGAACT 106
      |||:::||||| ||||| ||||| ||||| ||||| ||||| |||||
      53 lleuIleasnSerGlnTrpValIleuSerAlaAlaHisCysTyrLysSera 70
      107 CCTACACCATCGGCGTGGGCTGCACACAGTCTTGAGGCCGACCAAGACCA 156
      :::: ||||| ||||| ::::: ||||| |||||
      70 rgIleGlnValArgLeuGlyGlnTyrAsnIleAspVal..GlnGluasp 85
      157 GGGAGCCAGATGTTGGAGGCCCGACCTCTCCGTACGGCACCACCAAGATACAA 206

```

```

      86 serGIuValValArgSerSerSerValIleIleArgHisProLysTyrSe 102
      207 CAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG 256
      :   ::   |||  |||||:::|||||:::|||||:::|||||:::|||||
      102 rSerIleThrLeuAsnAsnAspIleMetLeuIleLysLeuAlaSerAlaVal 119
      257 TGNCCGAGTCTGCACACCATCCGAGCATCAGCATTCGCTTCGCAATGCCCT 306
      ||:::  |||  ||:::  ||:::  ||:::  ||:::  ||:::  |||
      119 alGIuTyrSerAlaAspIleGlnProIleAlaLeuProSerSerCysAla 135
      307 ACCGGGGGGAAGCTCTTGCCCTCGTNTCTGGCTGGGT...CTGCTGGCGAA 353
      |||||:::  |||||:::  |||||:::  |||||:::  |||||:::  |||||
      136 LysAlaGlyThrGluCysLeuIleSerGlyTyrGlyAsnThrLeuSerAs 152
      354 CGGC..AGATGCCCTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
      |||  ::  |||  ::|||::|||::|||::|||::|||::|||::|||
      152 nGIyTyrAsnTyrProGluLeuLeuGlnCysLeuAsnAlaProIleLeuS 169
      401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACACCCAGCATG 450
      ||:::  ||:::  ||:::  |||  ::|||
      169 eRAspGlnGluCysGlnGlnAlaTyrProGlyAspIleThrSerAsnMet 185
      451 TTCTGCGCGCGGCGGAGGCGACAGACCAAGAGACTCCTGCAACGCTGACTC 500
      ::|||::|||  ::  ::  |||||::|||::|||::|||::|||
      186 IleCysValGlyPheLeuGlnGlyGlyLysAspSerCysGlnGlyAspSe 202
      501 TGGGGGGCCCCGTGATCTGCAACGGGTACTTGACAGGGCCCTGTGCTTTG 550
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
      202 rGIyGlyProValValCysAsnGlyGluLeuGlnGlyIleValSerTyrp 219
      551 GAAAGCCCGGTGTGGCCAACTGGCGTGCAGGTGTCTACACCAACCTC 600
      ||  ||  ||:::  |||  |||||::|||::|||::|||::|||
      219 Iy...IleGlyCysAlaLeuLysGlyTyrProGlyValTyrThrLysVal 234
      601 TGCAAAATTCAGTGAAGTGAAGAAACCGTC 633
      ||:::  ||:::  |||||::|||::|||::|||::|||
      235 CysAsnTyrValAspTyrIleGlnGluThrIle 245
      seq_name: SwissProt_37:KLK8_RAT
      seq_documentation_block:
      ID_KLK8_RAT  STANDARD;  PRT;  261 AA.
      AC  P36374;
      DT  01-JUN-1994 (REL. 29, CREATED)
      DT  01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
      DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
      DE  GLANDULAR KALLIKREIN 8, PROSTATIC PRECURSOR (EC 3.4.21.35) (TISSUE
      KALLIKREIN) (P1 KALLIKREIN) (RGK-8).
      GN  KLK8 OR KLK-8.
      OS  RATTUS NORVEGICUS (RAT).
      OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
      RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
      RN  [1]
      RP  SEQUENCE FROM N.A.
      RX  MEDLINE; 89352606.
      RA  BRADY J.M., WINES D.R., MACDONALD R.J.;
      RT  "Expression of two kallikrein gene family members in the rat
      prostate.";
      RL  BIOCHEMISTRY 28:5203-5210(1989).
      CC  -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
      IN KININOGEN TO RELEASE LYS-BRADYKININ.
      CC  -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
      SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
      KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
      MET-|-XAA OR LEU-|-XAA.
      CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
      TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
      CC  -----
      CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
      CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
      CC  the European Bioinformatics Institute. There are no restrictions on its
      CC  use by non-profit institutions as long as its content is in no way

```


| | | | |
|-----|--------------------------------|-----------------------------|-----|
| 7 | CCCTGGCAGGCGGCACTGGTCTATGGA | AAACGAATGTGTTGCTGGGCGCT | 56 |
| | | | |
| 37 | ProTyrGlnValSerLeuAsnAlaGlySer | HisIle..CysGlyGly | 52 |
| 57 | CCTGGTGCATCCGCACTGGGTGCTGTCA | GGCGCACACACTGTTTCCAGAACT | 106 |
| | | | |
| 52 | rIeuIleThrAspGlnTrpValLeuSer | AlaAlaHisCysTyrHisProg | 69 |
| 107 | CCTACACCATCGGGGTGGGCTTGCA | CAGTCTTGAGGCGGCAAGACCA | 156 |
| | | | |
| 69 | IndeuGlnValArgLeuGlyGluHis | AsnIle..TyrGlnIleGly | 84 |
| 157 | GGGAGCCAGATGGTGGAGGCCAGCCT | CTCCGTACGGCACCCAGAGTACAA | 206 |
| | | | |
| 85 | AlaGluGlnPheIleAspAlaAlaLys | MetIleLeuHisProAspTyrAs | 101 |
| 207 | CAGACTCTTGCTCGCTAACCACTCAT | GCTCATCAAGTTGGAGGAATCCG | 256 |
| | | | |
| 101 | PlsTyrTrpThrValAspAsnAspIle | MetLeuIleLysLeuLysSerProA | 118 |
| 257 | TGTCCGAGTCTGACACCACTCGGAG | CATCAGCATTGCTTGCAGTGGCCCT | 306 |
| | | | |
| 118 | IaThrLeuAsnSerLysValSerThr | IleProLeuProGlnTyrCysPro | 134 |
| 307 | ACCGCGGGAACTCTTGCCCTCGTNT | CTGGCTGGGCTGCTGCGGCAACGG | 356 |
| | | | |
| 135 | ThrAlaGlyThrGluCysLeuValSer | GlyTyrGlyValLeuLysPheGln | 151 |
| 357 | C...AGAATGCCTACCGTGCCTGC | CACTGCGGTGAACGTGTGCGTGTCTG | 403 |
| | | | |
| 151 | YrPheGluSerProSerValLeuGln | CysLeuAspAlaProValLeuSerA | 168 |
| 404 | AGGANGTCTGCAGTAAGCTCTATGA | CCCGCTGTACCACCCCAGACATGTC | 453 |
| | | | |
| 168 | spSerValCysHisLysAlaIleTyr | ProArgGlnIleThrAsnAsnMetPhe | 184 |
| 454 | TGCGCGCGGCGGAGGCAAGACAGA | AGGACTCCTGCAACGGGTGACTGTGG | 503 |
| | | | |
| 185 | CysLeuGlyPheLeuGluGlyGlyLys | AspSerCysGlnTyrAspSerGln | 201 |
| 504 | GGGCGCCCTGATCTGCAACGGGTACT | TGCAAGGCGCTGTGTCTTTCGGAA | 553 |
| | | | |
| 201 | YcIlyProValValCysAsnGlyGlu | ValGlnGlyIleValSerTrpGlyA | 218 |
| 554 | AAGCCCCGTGTGGCCAACCTTGCG | GTGCCAGGTGTCTACACCAACCTCTGC | 603 |
| | | | |
| 218 | spGly...CysAlaLeuGluGlyLys | ProGlyValIleTyrThrLysValCys | 233 |
| 604 | AAATTCACCTGAGTGCATAGAGAAA | CCGTC 633 | |
| | | | |
| 234 | AsnTyrLeuAsnTrpIleGlnGlnThr | Val 243 | |

seq_name: SwissProt_37:KLK1_MACEA

seq_documentation_block:

| ID | KLK1_MACFA | STANDARD; | PRT; | 257 AA. |
|----|------------|-----------|------|---------|
|----|------------|-----------|------|---------|

AC Q07276;

DT 01-OCT-1994 (REL. 30, CREATED)

DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)

DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN).

GN KLK1. (GNAB FALTYA WADOTTA) (GUYWOF AND WOFWET) OS WADJCA FASCTOTIT ABTC

MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY).
 EIKAPKOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: PRIMATES.

OC EUNAKRIOIA; MEIAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES: CATARRHINI: CERCOPITHECINAE: MACACA

УЧЕНИКИ; КАЛАНД
[1]

RP SENTENCE FROM N. A.

RX MEDLINE: 93305727

BA LIN F. K., LIN C. H., CHOU C. C., CHEN K. C., TII H. S., BACHTELIER B.

RA HERRERA C., JONES T., CHAO J., CHAO L.:

RT "Molecular cloning and sequence analysis of the monkey and human

RT, -tissua kallikrein genes. ";

```

RL BIOCIM. BIOPHYS. ACTA 1173:325-328(1993).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L10039; G293141; -.
DR PIR; S33772; S33772.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00752; 1H1A.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS;
KW ZMOGEN; SIGNAL.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 257 GLANDULAR KALLIKREIN 1.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
FT CARBOHYD 101 101 POTENTIAL.
FT CARBOHYD 105 105 POTENTIAL.
FT CARBOHYD 160 160 POTENTIAL.
FT CARBOHYD 162 162 POTENTIAL.
FT DISULFID 31 169 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 148 215 BY SIMILARITY.
FT DISULFID 180 194 BY SIMILARITY.
FT DISULFID 205 230 BY SIMILARITY.
SQ SEQUENCE 257 AA; 28237 MW; 82A709DC CRC32;

```

alignment_scores:

Quality: 461.50

Percent Siml

alignment_block:

US-09-030-606-175 x KLK1_MACFA

Align seg 1/1 to: KLK1_MACFA from: 1 to: 257

```

1  GCGCAGCCCTGGCAGGGCGGCACTGGTCATGGAAGAAGATTGTTGCTC  50
   :::::::::::::::::::::::::::: :::: :::: ::::
32  SerGlnProTTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysG1  48
51  GGGCGTCTGTGGTCATCCGCGCAGTGGGTGCTGTCAAGCCGCACACTGTTCC  100
   :::::::::::::::::::::::::::::: :::: :::: ::::
48  YGlyIleLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIles  65
101  AGAACTCCTAACCCATCGGGGCTGGGCTGCACACAGTCTTGAGCCGACCA  150
   :::::::::: :::: :::: :::: :::: ::::
65  erAspasnIyGlnLeuTrpLeuGlyArgHisasnLeu..PheasnAsp  80
151  GAGCCAGGGAGCCAGATGTTGGAGGCCAGCCTTCCTACGGCACCACAGA  200
   ||| :::: :::: :::: :::: ||| |||||
81  GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProG1  97
201  GTACACACAGACTCTTGCTCGCTAAC.....  225

```



```

117 luserAsnAspleuMetleuLeuAArgleuSerGluProAlaAspleThr 133
268 GACACCATCCGGAGCATCAGCATGTGCTTCGCAGATGCCCTACCGCGGGAA 317
134 AspalavallysProileAlaLeuProThrGluGluProLysleuGlySe 150
318 CTCTGGCCCTGNTCTGGCTGGGCTGCTGCTGGCGAAGGAGATG... 363
150 rThCysleuValserGlyTrpGlySerIleThrProThrLysPheGlnT 167
364 .CCTACCGTGTGCTGCTGCTGTAACGTGTGCGTGTGCTGAGGANGTC 411
167 hrProAspAspleuGlnCysValSerIleLysleuLeuProAsnGluVal 183
412 TGCAGTAAGCTCTATGACCCGCTGTACACCCACAGCATGTCTGCCCGG 461
184 CysValLysAsnHisAsnGlnLysValThrAspValMetLeuCysAlaG1 200
462 CGGAGGGCAAGACCAAGAGACTCCTGCAAGGTGACTCTGGGGGCCCC 511
200 yGluMetGlyGlyGlyLysAspThrCysLysGlyAspSerGlyGlyProL 217
512 TGATCTGCAACGGGTACTTGCAGGGGCTGTGTCTTTGGAAAAAGCCCC 561
217 euileCysaspGlyValLeuHisGlyIleThrAlaTrpGlyProilePro 233
562 TGTGGCCAACCTTGGCGTGCAGGTGTCTACACCAACCTCTGCAATTCAC 611
234 CysGlyLysProAsnThrProGlyValTyrThrLysleuIleLysPheTh 250
612 TGAGTGATAGAGAAACCGTC 633
250 rAsnTrpIleLysAspThrMet 257
```


This Page Blank (uspto)

OM of: US-09-030-606-175 to: SPTREMBL_10:* out_format : pfs
Date: Sep 25, 1999 10:35 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL-frame+n2p.model -DEV-rlp
-Q/cgn2_1/USPTO_spool/US09030606/runat.24091999_171617_29869/app-query.fasta.1
-DB-SPTREMBL_10 -QFMT-fastan -SUFFIX-rspt -GAPOP-12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-biosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE-LOCAL -OUTFMT-pfs -NORM=stat -USER-US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-175
Query length: 1167
Database: SPTREMBL_10:*
Database sequences: 201082
Database length: 61543640
Search time (sec): 260.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|----------------------|------|--------|---------|---------|-----|-----------------------------------|
| sp_rodent:Q920M1 | + | 834.00 | 1336.97 | 5.5e-67 | 255 | Q920M1 mus musculus (mouse). en |
| sp_rodent:Q63274 | + | 486.00 | 776.26 | 1.0e-35 | 235 | Q63274 rattus norvegicus (rat). s |
| sp_human:Q75837 | + | 485.50 | 773.87 | 1.1e-35 | 282 | Q75837 homo sapiens (human). se |
| sp_rodent:Q63275 | + | 477.00 | 761.59 | 6.5e-35 | 239 | Q63275 rattus norvegicus (rat). s |
| sp_mammal:Q29474 | + | 466.00 | 743.08 | 6.4e-34 | 261 | Q29474 canis familiaris (dog). l |
| sp_rodent:Q88309 | + | 462.00 | 736.63 | 1.5e-33 | 261 | Q88309 mus musculus (mouse). pr |
| sp_rodent:Q54854 | + | 454.50 | 724.87 | 6.9e-33 | 251 | Q54854 rattus norvegicus (rat). s |
| sp_vertebrate:Q93265 | + | 448.50 | 715.22 | 2.4e-32 | 250 | Q93265 pleuroctes americanus. p |
| sp_vertebrate:Q92046 | + | 447.00 | 712.84 | 3.3e-32 | 249 | Q92046 dissostichus mawsoni. pr |
| sp_rodent:Q61855 | + | 447.00 | 712.43 | 3.3e-32 | 261 | Q61855 mus musculus (mouse). tl |
| sp_rodent:Q88301 | + | 441.50 | 704.07 | 1.0e-31 | 246 | Q88301 mus musculus (mouse). se |
| sp_vertebrate:Q93266 | + | 438.50 | 699.37 | 1.9e-31 | 242 | Q93266 pleuroctes americanus. s |
| sp_vertebrate:Q42608 | + | 434.00 | 691.94 | 4.8e-31 | 247 | Q42608 petromyzon marinus (sea |
| sp_vertebrate:Q42158 | + | 434.00 | 691.94 | 4.8e-31 | 247 | Q42158 petromyzon marinus (sea |
| sp_rodent:Q921R9 | + | 431.50 | 687.94 | 8.1e-31 | 246 | Q921R9 mus musculus (mouse). tl |
| sp_vertebrate:Q91515 | + | 430.50 | 686.65 | 9.9e-31 | 237 | Q91515 fuigu rubripes (japanese |
| sp_vertebrate:Q92099 | + | 430.00 | 685.66 | 1.1e-30 | 242 | Q92099 paranothemia magellani |
| sp_vertebrate:Q42160 | + | 426.00 | 679.10 | 2.5e-30 | 245 | Q42160 petromyzon marinus (sea |
| sp_human:Q15665 | + | 420.50 | 669.74 | 7.9e-30 | 259 | Q15665 homo sapiens (human). tl |
| sp_vertebrate:Q42159 | + | 420.00 | 669.46 | 8.7e-30 | 244 | Q42159 petromyzon marinus (sea |
| sp_vertebrate:Q9YGS1 | + | 416.00 | 663.37 | 2.0e-29 | 234 | Q9YGS1 agkistrodon acutus (dein |
| sp_vertebrate:Q13059 | + | 402.50 | 640.74 | 3.3e-28 | 258 | Q13059 trimeresurus gramineus (|
| sp_vertebrate:Q9YGS8 | + | 398.50 | 633.60 | 7.5e-28 | 279 | Q9YGS8 gloydius halys brevicaud |
| sp_rodent:Q03955 | + | 398.00 | 633.75 | 8.3e-28 | 250 | Q03955 promys natalensis (afri |
| sp_vertebrate:Q91516 | + | 382.50 | 608.47 | 2.0e-26 | 258 | Q91516 trimeresurus stejnegeri |
| sp_vertebrate:Q91053 | + | 378.50 | 601.89 | 4.7e-26 | 262 | Q91053 gloydius ussuriensis. ca |
| sp_vertebrate:Q9YGI6 | + | 376.50 | 598.73 | 7.1e-26 | 260 | Q9YGI6 agkistrodon halys pallas |
| sp_vertebrate:Q9YGS2 | + | 375.50 | 597.11 | 8.7e-26 | 260 | Q9YGS2 gloydius halys (halys vl |
| sp_vertebrate:Q13069 | + | 374.50 | 595.60 | 1.1e-25 | 257 | Q13069 bothrops jararaca (jarar |
| sp_vertebrate:Q91510 | + | 371.50 | 590.76 | 2.0e-25 | 257 | Q91510 trimeresurus mucroscuam |
| sp_vertebrate:Q93421 | + | 370.50 | 589.85 | 2.4e-25 | 237 | Q93421 agkistrodon halys pallas |
| sp_vertebrate:Q91509 | + | 370.50 | 589.15 | 2.5e-25 | 257 | Q91509 trimeresurus mucroscuam |
| sp_vertebrate:Q13061 | + | 369.50 | 587.82 | 3.0e-25 | 260 | Q13061 trimeresurus gramineus (|
| sp_vertebrate:Q91508 | + | 368.50 | 585.43 | 3.7e-25 | 260 | Q91508 trimeresurus gramineus (|
| sp_vertebrate:Q91500 | + | 367.50 | 584.31 | 4.6e-25 | 257 | Q91500 trimeresurus mucroscuam |
| sp_vertebrate:Q42207 | + | 366.50 | 582.69 | 5.6e-25 | 257 | Q91500 trimeresurus mucroscuam |
| sp_vertebrate:Q91036 | + | 364.00 | 581.85 | 6.2e-25 | 258 | Q42207 gloydius ussuriensis. ca |
| sp_vertebrate:Q91511 | + | 363.50 | 580.05 | 9.2e-25 | 219 | Q91036 gadus morhua (atlantic c |
| sp_vertebrate:Q93502 | + | 363.50 | 577.75 | 1.0e-24 | 257 | Q91511 trimeresurus mucroscuam |
| sp_vertebrate:Q73800 | + | 362.50 | 576.14 | 1.3e-24 | 260 | Q93502 gloydius halys (halys vl |
| sp_vertebrate:Q13057 | + | 362.50 | 576.14 | 1.3e-24 | 260 | Q73800 gloydius halys (halys vl |

sp_vertebrate:Q91507 + 361.50 574.63 1.6e-24 257 | Q91507 trimeresurus mucroscu
sp_vertebrate:Q13063 + 361.50 574.59 1.6e-24 258 | Q13063 trimeresurus gramineu
sp_vertebrate:Q13062 + 360.50 573.01 1.9e-24 257 | Q13062 trimeresurus gramineu

seq_name: sp_rodent:Q920M1

seq_documentation_block:
ID Q920M1 PRELIMINARY; PRT; 255 AA.
AC Q920M1;
DT 01-MAY-1999 (TREMBLREL. 10, Created)
DT 01-MAY-1999 (TREMBLREL. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLREL. 10, Last annotation update)
DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SWISS-WEBSTER;
RA Stimmer J.;
RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF019979; AAC98894.1; -
KW Signal.
FT SIGNAL. 1 25 POTENTIAL.
FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.
SQ SEQUENCE 255 AA; 27488 MW; 6E711616 CRC32;

alignment_scores:
Quality: 834.00 Length: 214
Ratio: 4.366 Gaps: 0
Percent Similarity: 89.252 Percent Identity: 68.692

alignment_block:
US-09-030-606-175 x Q920M1 ..

Align seg 1/1 to: Q920M1 from: 1 to: 255

1 GCGCAGCCCTGGCAGGGCGGCGACTGTCATGGAAACGAATTGTTCTGCTC 50
:::|||||
42 SerGlnProTyrGlnAlaAlaLeuPheSerGlnAspGlyPhePheCysSe 58
51 GGGCGTCTGGTGCATCCGCGAGTGGTGTGTCAGCCGACACTGTTCC 100
|||||
58 rGlyValLeuValHisProGlnTyrValLeuSerAlaAlaHisCysLeuG 75
101 AGACTCCCTACACCATCGGGCTGGGCTGCACAGTCTGAGGCCGACCAA 150
|||||
75 lngluserTyrIleValGlyLeuGlyLeuHisAsnLeuGlySerGln 91
151 GAGCCAGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGCACCAGAGA 200
|||||
92 GluProGlySerArgMetLeuGlnAlaHisLeuSerIleGlnHisProAs 108
201 GTACACAGACTCTTGTGCTGCTAACGACCTCATGCTCATCAAGTTGGACG 250
:::|
108 nPheAsnSprProSerPheAlaAsnAspLeuMetLeuIleLysLeuAsnG 125
251 AATCCGTCGTCAGTCTGACACACATCCGAGCATCAGCATTCGTCGACG 300
|||||
125 luserValIleGlnSerAsnThrIleArgSerIleProValAlaThrGln 141
301 TGCCCTACGGCGGGAACCTCTGCGCTGTCGTCGTCGTCGTCGTCGTCG 350
|||||
142 CysProThrProGlyAspThrCysLeuValSerGlyTyrPglGlnLeuLys 158
351 GAACGGCAATGCTACCTACCTGCTGCACCTGCGTGAACGTCGTCGTCG 400
|||||
158 sAsnGlyLysLeuProSerLeuGlnCysValAsnLeuSerValAlaAs 175
401 CTGAGGANGTCTGACAGTAAGCTCTATGACCCGCTGTACACCCAGCATG 450
|||||


```
175 erGlutIurHrcYsarGleuLeuTyraSProValTyRHisleuSerMet 191
451 TTCTGCGCCCGCGGAGGCAAGACCAAGAGACTCCTGCAACGGTGACTC 500
|||||
192 PheCysAlaGlyGlyGlyGlnAspGlnLysAspSerCysAsnGlyAspSe 208
501 TGGGGGGCCCCCTGATCTGCAACGGGTACTTGACGGGCTTGCTCTTCG 550
|||||
208 rGlyGlyProIleValCysAsnArgSerLeuGlnGlyLeuValSerMetG 225
551 GAAAAGCCCCGTGTGGCCAACCTGGCGTGCCAGGTGTCTACACCACTC 600
|||||
225 LyGlnGlyLysCysGlyGlnProGlyIleProSerValTyThrAsnLeu 241
601 TGCAAAATCACTGAGTGGATAGAGAAAACCGTCCAGNCCAGT 642
|||||
242 CysLysPheThrAsnTrpIleGlnThrIleIleGlnThrAsn 255
seq_name: sp_rodent:Q63274
```

```
seq_documentation_block:
ID Q63274 PRELIMINARY; PRT; 235 AA.
AC Q63274;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; I33839; AAA58781.1; -.
DR PFAM; PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 235 AA; 26226 MW; E6DE5AD8 CRC32;
```

```
alignment_scores:
Quality: 486.00 Length: 227
Ratio: 3.019 Gaps: 5
Percent Similarity: 70.925 Percent Identity: 42.291
```

alignment_block:
US-09-030-606-175 x Q63274 ..

Align seg 1/1 to: Q63274 from: 1 to: 235

```
1 GCGCAGCCCTGGCAGCGGCACTGTCATGGAAGAATTGTTCTGCTC 50
|||||
11 SerGlnProTrpGlnValAlaValIle.....AsnArgTyrlaucysgl 25
51 GGGCGTCTGTGCATCCGCAAGTGGTGTCTGCAGCCGCACTGTTCC 100
|||||
25 yGlyValLeuIleAspProSerTrpValIleThrAlaAlaHisCysTyrS 42
101 AGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAA 150
|||||
42 erHisTyTrpHisValLeuLeuGlyArgHisAsnLeuPheGluAsp... 57
151 GAGCCAGGAGCCAGATGGTGAGGCGAGCCCTCTCCGTACGCGACCCAGA 200
|||||
58 GluProPheAlaGlnTyArgPheValSerGlnSerPheProHisProAs 74
201 GTACAACAGACTCTTGCTC..... 219
|||||
74 pTyraSnProPheLeuMetArgAsnHisThrArgGlnThrGlyTyraSPt 91
220 ..GCTAACGACTCATGTCATCATCAAGTTGGACGAATCCGTGCCAGTCT 267
|||||
```

```
91 yrSerAsnAspLeuMetLeuLeuHisLeuSerGluProAlaAspIleThr 107
268 GACACCATCCGAGACATCAGCATTGCTTCGCACTGCCCTACCCGGGGAA 317
|||||
108 AspGlyValLysValIleAspLeuProThrGluGluProLysValGlySe 124
318 CTCTTGCCCTCGTNTCTGGCTGGGGT.....CTGCTGGCGAAGC 355
|||||
124 rThrcysLeuValSerGlyTrpGlySerThrLysProLeuIleSer... 139
356 GCAGATGCCTTACCGTGTGCACACTGCGTGAACGTGTGGTGTCTGAG 405
|||||
140 ..GluLeuProAspAspLeuGlnCysValAsnIleAspLeuLeuSerAsn 155
406 GANGTCTGCAGTAAAGCTGTATGACCCGCTGTACCAACCCAGCATGTCTG 455
|||||
156 GluLysCysIleGlnAlaTyraArgTrpLysValThrAspLeuMetLeuCy 172
456 CGCCGCGGAGGCAAGACCAAGAGACTCCTGCAACGGTGACTCTGGGG 505
|||||
172 salaglyLysLeuGlnGlyGlyLysAspAlaCysAsnGlyAspSerGlyG 189
506 GGCCTGATCTGCAACGGGTACTTGACAGGCGCTGTGTCTTTGCGAAA 555
|||||
189 LyProLeuIleCysAspGlyValLeuGlnGlyLeuThrSerTrpGlySer 205
556 GCCCCGTGTGGCCAACCTGGCGTGCAGGTGTCTACACCAACCTGTGCAA 605
|||||
206 ValProCysGlyGluProHisAsnProGlyIleTyThrLysIleIleLy 222
606 ATTCACGTGATGATAGAGAAAACCGTCCAG 636
|||||
222 sPheThrSerTrpIleLysGluValMetLys 232
seq_name: sp_human:O75837
```

```
seq_documentation_block:
ID O75837 PRELIMINARY; PRT; 282 AA.
AC O75837;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE SERINE PROTEASE (TLSP).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98438738.
RA YOSHIDA S., TANIGUCHI M., SUEMOTO T., OKA T., HE X.P., SHIOSAKA S.;
RT "CDNA cloning and expression of a novel serine protease, TLSP.";
RL Biochim. Biophys. Acta 1399:225-228(1998).
DR EMBL; AB012917; BAA33404.1; -.
DR PFAM; PF00089; trypsin; 1.
KW Protease; Serine protease.
FT CHAIN 54 282 TLSP.
SQ SEQUENCE 282 AA; 31059 MW; E97E294C CRC32;
```

```
alignment_scores:
Quality: 485.50 Length: 218
Ratio: 3.053 Gaps: 3
Percent Similarity: 72.936 Percent Identity: 41.284
```

alignment_block:
US-09-030-606-175 x O75837 ..

Align seg 1/1 to: O75837 from: 1 to: 282

```
1 GCGCAGCCCTGGCAGCGGCACTGTCATGGAAGAATTGTTCTGCTC 50
|||||
64 SerGlnProTrpGlnAlaAlaLeuPheGluLysThrArgLeuLeuCySgl 80
|||||
```



```

51 GGGCGCTCTGGTGCATCCGCGAGTGGGTGCTGCAGCCGACACATGTTTCC 100
   ::::::::::: ||:::||||:||||:||||:||||:||||:
80 yAlaThrLeuIleAlaProArgTrpLeuLeuThrAlaAlaHisCysLeuL 97
   :: ||| :::: ||||| ||:::||||: ::::
101 AGAAGCTCTACACCATCGGGGTGGGCTGCACAGTCTTGAGGGCGACCAA 150
   :: ||| :::: ||||| ||:::||||: ::::
97 ySProArgTrpLeuIleValHisLeuGlyGlnHisAsnLeuGln..LysGlu 112
   ||| :::: ||||| ||::: ||| |||||
113 GluGlyCysGluGlnInThrArgThrAlaThrGluSerPheProHisProGl 129
   ||| :::: ||||| ||::: ||| |||||
201 GTACAACAGACTCTTG.....CTCGCTAACGACCTCATGCTCA 238
   ::::: ||| |||||:|||||:
129 yPheAsnAsnSerLeuProAsnLysAspHisArgAsnAspIleMetLeuV 146
   ::::: ||| |||||:|||||:
239 TCACAGTTGGACGAATCCGTGTCCGAGTCTGCACACCATCCGGAGCATCAGC 288
   ||::: ||::: ||||| ::::: ||| ::::: |||
146 aLysMetAlaSerProValSerIleThrTrpAlaValArgProLeuThr 162
   ||::: ||::: ||||| ::::: ||| ::::: |||
289 ATTGCTTCGCAGTGCCTTACCGCGGGGAACTCTTGCCCTCGTNTTGCGTG 338
   ::::: ||::: ||| |||||:|||||:|||||:|||||
163 LeuSerSerArgCysValThrAlaGlyThrSerCysLeuIleSerGlyTr 179
   ::::: ||::: ||| |||||:|||||:|||||:|||||
339 GGGTCTGCTGGCGAAGCGC....AGAATGCCCTACCGTGCCTGCATCGG 382
   |||| ::::: ||::: ||| ::::: ||| ::::: |||
179 pGlySerThrSerSerProGlnLeuArgLeuProHisThrLeuArgCysA 196
   |||| ::::: ||::: ||| ::::: ||| ::::: |||
383 TGCACGTGTCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCG 432
   ::||:::||||:||||: ||::: |||
196 IaAsnIleThrIleIleGlnHisGlnLysCysGluAsnAlaTrpProGly 212
   ::||:::||||:||||: ||::: |||
433 CTGTACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAAGA 482
   :::: ||| |||||: ||::: |||
213 AsnIleThrAspThrMetValCysAlaSerValGlnGluGlyLysAs 229
   :::: ||| |||||: ||::: |||
483 CTCCTGCACACGGTGACTCTGGGGGCGCCCTGATCTGCACAGGGTACTTGC 532
   |||||:||||:||||:||||:||||: ||||
229 pSerCysGlnGlyAspSerGlyGlyProLeuValCysAsnGlnSerLeug 246
   |||||:||||:||||:||||:||||: ||||
533 AGGGCCTTGTTCTTTCGGAAGAACCCCGTGTGGCCAACTTGGCGTCCA 582
   |||||:||||:||||:||||: |||||: |||
246 InGlyIleIleSerTrpGlyGlnsPProCysAlaIleThrArgLysPro 262
   |||||:||||:||||:||||: |||||
583 GGTGTCTACACCAACCTCTGCAAATTCACGTAGTGGATAGAGAAACCGT 632
   |||||:||||:||||:||||:||||: |||||: |||
263 GlyValIleTrpThrLysValCysLysTyrValAspTrpIleGlnIuThrMe 279
   |||||:||||:||||:||||:||||: |||||: |||
633 CCAG 636
   ::::
279 tLys 280

```

```

seq_name: sp_rodent:Q63275
seq_documentation_block:
ID Q63275 PRELIMINARY; PRT: 239 AA.
AC Q63275:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U33840; AAA58782.1; -.
DR PFIAM: PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 239 AA; 26382 MW; 0609E3E6 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 477.00 | Length: 224 |
| Ratio: | 2.944 | Gaps: 5 |
| Percent Similarity: | 72.321 | Percent Identity: 41.518 |

alignment_block:

US-09-030-606-175 x Q63275

Align seg 1/1 to: Q63275 from: 1 to: 239

```

1  GCGCAGCCCTGGCAGCGCGGCACCTGGTCTATGGAAACGAATTGTTGGCTC 50
   :::::::::::::::::::::::::::: ||||| :::::
19  SerGlnProTrpGlnValAlaValIle.....AsnGluAspLeuGlySerG1 33
51  GGGCGTCTGGTGGATCCGCGAGTGGGTGGTGTCTGACCGGCACACTGTGTCC 100
   :::::::::::::: ||:::|||||:|||||:|||||:
33  GcLValLeuIleAspProSerTrpValIleThrAlaAlaIleAsnGlySerG1 50
101  AGAAGCTCTACACCAATCGGGCTGGGCTGGCAGCATGTTGTAGGCGGCA 150
   :::::||||| :::: ||||| :::::|||||: |||
50  eAspAsnTrpThrHisValLeuLeuGlnGlnAsnAsnLeuSerGlnAspVal 66
151  GAGCCAGGGAGGCGCAGATGGTGGAGGCGCAGCCTCTCCGTACGGCAGCCAGA 200
   :::: :::::||||| ||| ||| |||||:
67  Gln.....HisArgLeuVal.....SerGlnSerPheArgHisProAs 79
201  GTACACAGACTTTGCTCTC.....G 220
   ::::: :::::|||||:
79  rTyrLysProPheLeuMetArgAsnHisThrArgLysProLysAspTyrS 96
221  CTAAAGCACTCATGCTCATCAAGTTGAGCAATCCGTTGTCGAGTCTGAC 270
   ::|||:|||||:|||||: ||:::||||| ::::: |||
96  eAsnAspLeuMetLeuLeuHisLeuSerGlnProAlaAspIleThrAsp 112
271  ACCATCCGGAGCATCAGCATGTGCTTCCGACATCCCTACCGCGGCGAATC 320
   ::::: ||:::|||||: ::::: ||| :::|||||:
113  GcLValValValIleAspLeuProThrLysGlnProLysValGlySerTh 129
321  TTGCCTCGTNTCTGGCTGGGCTCTGCTG.....GCGAACGGCAGAAATGC 364
   ::|||:|||||:|||||:|||||: ::::: |||:
129  rCysLeuValSerGlyTrpGlySerThrAsnProSerGluTrpGlnPheP 146
365  CTACCGTGTGCTGCACTGCGTGAACGTGCGGTGGTGTCTGAGANGTCTGC 414
   || |||:|||||:|||||: :::::|||||: |||
146  roAspAspLeuGlnCysValAsnIleHisLeuLeuSerAsnGlnLysCys 162
415  AGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGGCGCGCGG 464
   ||| ||| |||:|||||:
163  IleLysAlaIleTyrLysGlnLysValThrAspLeuMetLeuCysAlaGlyG1 179
465  AGGGCAGAGCCAGAGGACTCCTGCAACGSGTGACTTGGGGGGCCCTGA 514
   :::: |||||:|||||:|||||:|||||:|||||
179  uLeuGlnGlyGlyLysAspThrCysArgGlyAspSerGlyGlyProLeuI 196
515  TCTGCACAGGGTACTTGCAGGGCCTGTGTCTTTCGGAAAGGCCCGGTGT 564
   |||||:||||| |||||:|||||:|||||:|||||
196  leCysAspGlyValLeuGlnGlyIleThrSerTrpGlySerValProCys 212
565  GGCCCAACTTGGCGTGGCCAGGTGTGTACCAACCAACTCTGCAAAATTCAC 614
   |||: :::: |||||:|||||:|||||: ||| |||||:
213  GcLysLysProAsnLysProGlyIleTyrThrLysLeuIleLysPheThrSe 229
615  GTGGATAGAGAAAACCGTCCAG 636
229  rTrpIleLysGlnValMetLys 236

seq_name: sp_mammal:Q29474
seq_documentation_block:
ID  Q29474 PRELIMINARY; PRT; 261 AA.
AC  Q29474;

```


DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)
DE KALLIKREIN PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (GLANDULAR KALLIKREIN) (KININOGENIN).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MONGREL;
RX MEDLINE; 94250683.
RA GAUTHIER E.R., DUMAS C., CHAPDELAIN P., TREMBLAY R.R., DUBE J.Y.;
RT "Characterization of canine pancreas kallikrein cDNA.";
RL Biochim. Biophys. Acta 1218:102-104(1994).
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-I-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES.
CC HIGHLY SELECTIVE ACTION TO RELEASE KALLIDIN
CC (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-I-XAA OR LEU-I-XAA.
CC THE RAT ENZYME IS UNUSUAL IN LIBERATING BRADYKININ DIRECTLY FROM
CC AUTOLOGOUS KININOGENS BY CLEAVAGE AT TWO ARG-I-XAA BONDS.
DR EMBL; X75479; CAA53210.1; -.
DR PRAM; PF00089; trypsin; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 261 KALLIKREIN.
SO SEQUENCE 261 AA; 28913 MW; 65B49AEF CRC32;

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 466.00 | Length: 224 |
| Ratio: | 2.913 | Gaps: 3 |
| Percent Similarity: | 71.429 | Percent Identity: 39.732 |

```
alignment_block:
US-09-030-606-175 x Q29474 .
```

Align seg 1/1 to: Q29474 from: 1 to: 261

[illegible]

```

167 yrrProaspaSpLeuGlncIncysvalaSpLeuGlulLeuLeuSerAsnaspIle 183
167 yrrProaspaSpLeuGlncIncysvalaSpLeuGlulLeuLeuSerAsnaspIle 183
412 TGCAGTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTGGCGCG 461
412 TGCAGTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTGGCGCG 461
184 CysAlaAsnAlaHisSerGlnLysValThrGluPheMetLeuCysAlaG1 200
184 CysAlaAsnAlaHisSerGlnLysValThrGluPheMetLeuCysAlaG1 200
462 CGGAGGGCAAGACAGAAAGGACTCCTGCAACGGTGACTCTGGGGGGCCC 511
462 CGGAGGGCAAGACAGAAAGGACTCCTGCAACGGTGACTCTGGGGGGCCC 511
200 yHisLeuGlulGlyGlyLysAspThrCysValGlyAspSerGlyGlyProl 217
200 yHisLeuGlulGlyGlyLysAspThrCysValGlyAspSerGlyGlyProl 217
512 TGATCTGCACACGGGTACTTGCAGGGCCCTGTGTCTTTCCGGAAGACCCC 561
512 TGATCTGCACACGGGTACTTGCAGGGCCCTGTGTCTTTCCGGAAGACCCC 561
217 eutLeCysaspGlyValLeuGlngLyleThrSerTrpGlyHisValPro 233
217 eutLeCysaspGlyValLeuGlngLyleThrSerTrpGlyHisValPro 233
562 TGTGGCCAACCTTGGCGGTGCCAGGTGTCTACACCAACCTCTGCAAAATTAC 611
562 TGTGGCCAACCTTGGCGGTGCCAGGTGTCTACACCAACCTCTGCAAAATTAC 611
234 CysGlySerProAsnMetProAlaValTyThrLysValIleSerHisIle 250
234 CysGlySerProAsnMetProAlaValTyThrLysValIleSerHisIle 250
612 TGAGTGGATAGAGAAAACCGTC 633
612 TGAGTGGATAGAGAAAACCGTC 633
250 uGluTrpIleLysGluThrMet 257
250 uGluTrpIleLysGluThrMet 257

```

seq_name: sp_rodent:088309

```
seq_documentation_block:
```

DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
DE PRORENIN-CONVERTING ENZYME (MK13B) PRECURSOR (MK13B).

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98351995.
RA HOSOI K., TADA J., TSUMURA K., KANAMORI N., YAMANAKA N.;
RT "Expression of an allozyme of prorenin-converting enzyme in the
RT submandibular gland of DBA/2N mice.";
RL J. Biochem. 124:368-376(1998).
DR EMBL; AB016032; BAA31686.1; -.
DR PFM; PF00089; trypsin; 1.
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
SQ SEQUENCE 261 AA; 28692 MW; 8DB1814F CRC32;

```

alignment_scores:
  Quality: 462.00
  Ratio: 2.981
  Gaps: 3
Percent Similarity: 69.196
Percent Identity: 41.071

```

```
alignment_block:
US-09-030-606-175 x 088309 ..
```

Align seg 1/1 to: 088309 from: 1 to: 261

```

1  GCGCAGCCCTGGCAGGGCGGCACSTGCTCATGGAAACGAATTGTGTGCTC 50
   :::::::::::::::::::::  ::::::::::  :::::
35  SerGlnProTrpGlnValAlaValAlaTyrTyrGlnLysGlnHisIleCysG1 51
   :::::::::::::::::::::  ::::::::::  :::::
51  GGGCGCTCTGTGTCATCCGCAGATGGGTCTGTCAAGCGCACACTGTATCC 100
   :::::::::::::::::::::  ::::::::::  :::::
51  yGlnValLeuLeuAspArgAsnTrpValLeuThrAlaAlaHisCysTyrV 68
   :::::::::::::::::::::  ::::::::::  :::::
101  AGAATCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGGACCAA 150
   ::::::::::  ::  :::::  :::::  :::::
68  alaSpGlnTyrGlnValTrpLeuGlyLysAsnLysLeu..PheGlnGln 83
   ::::::::::  ::  :::::  :::::  :::::
151  GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGATACGGCACCCAGA 200

```



```
|||||:|||||:|||||:|||||
84 GluProSerAlaGlnHisArgLeuValSerLysSerPheProHisProG1 100
201 GTACACAGACTCTGCTC..... 219
|||||
100 YTyAsnMetSerLeuMetLeuGlnThrIleProProGlyAlaAsp 117
220 .GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGCTT 267
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 heserAsnAspLeuMetLeuArgLeuSerLysProAlaAspIleThr 133
268 GACACCATCCGAGACATTCAGCATTGCTTCGACGTGCCCTACCGCGGGA 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 AspValValLysProIleAlaLeuProThrLysGluProLysProGlySe 150
318 CTCTTGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGA.....A 361
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 rLysCysLeuAlaSerGlyTrpGlySerIleThrProThrArgTrpGlnL 167
362 TGCCCTACCGTGCCTGCACCTCGTGAACGTGTCGGGTGCTGTGAGGANGTC 411
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ySProAspAspLeuGlnCysValPheIleThrLeuLeuProAsnGlnAsn 183
412 TGCAGTAGCTCTATGACCCGCTGTACACCCCAAGCATGTTCTGCGCCG 461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 CysAlaLysValTyrLeuGlnLysValThrAspValMetLeuCysAlaG1 200
462 CGGAGGCGAAGACCAAGAGCACTCCTGCACACGGTGAAGTCTGGGGGCC 511
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 yGluMetGlyGlyLysAspThrCysArgAspAspSerGlyGlyProL 217
512 TGATCTGCAACGGGCTACTTGACAGGGCTGTGTCTTTCGAAAGCCCC 561
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 euIleCysaspGlyIleLeuGlnGlyThrThrSerTyrGlyProThrPro 233
562 TGTGCCCAACTGGCGTGCAGAGGTGTCTACACCAACTCTGCAAAATTCAC 611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 CysGlyLysProGlyValProAlaIleTyrThrAsnLeuIleLysPheAs 250
612 TGAGTGGATAGAGAAACCGTC 633
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 nserTrpIleLysAspThrMet 257
```

seq_name: sp_rodent:054854

seq_documentation_block:

```
ID 054854 PRELIMINARY; PRT; 251 AA.
AC 054854;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE MYLENCEPHALON SPECIFIC PROTEASE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE; 97477435.
RA SCARISBRICK I.A., TOWNER M.D., ISACKSON P.J.;
RT "Nervous system-specific expression of a novel serine protease:
RT regulation in the adult rat spinal cord by excitotoxic injury.";
RL J. Neurosci. 17:8156-8168(1997).
DR EMBL; AF016269; AAC02300.1; -.
DR PFM; PF00089; trypsin; 1.
KW Protease.
SQ SEQUENCE 251 AA; 28014 MM; 21D0D8A9 CRC32;
```

alignment_scores:
Quality: 454.50 Length: 212
Ratio: 2.971 Gaps: 1
Percent Similarity: 72.170 Percent Identity: 39.151

alignment_block:
US-09-030-606-175 x 054854 ..

Align seg 1/1 to: 054854 from: 1 to: 251

```
1 GCGCAGCCCTGGCAGCGGCACACTGGTCATGGAAACGAATGTTCTGCTC 50
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 SerHisProPheGlnAlaAlaLeuTyrThrSerGlyHisLeuLeuCysG1 55
51 GGGCGTCTGTGTCATCCGACGTGGGTGCTGTACGCCGACACTGTTTC 100
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 yGlyValLeuValGlyProGlnTrpValLeuThrAlaAlaHisCysLysL 72
101 AGAACTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGCGCACCAA 150
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 ySProAsnLeuGlnValTyrLeuGlyLysHisAsnLeuArg...GlnThr 87
151 GAGCCAGGAGGACGAGATGGTGAGGCCACCTCTCCGTACGGCACCCAGA 200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 GluThrPheGlnArgGlnIleSerValAspArgThrIleValHisProAr 104
201 GTACACAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACG 250
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 GTyrAsnProGlnThrHisAspAsnAspIleMetValHisLeuLysA 121
251 AATCCGTGCCAGTCTGACACCATCCGAGCATCAGCATGCTTCCGACG 300
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 rGProValLysPheSerGlnArgIleGlnProLeuProLeuLysLysAsp 137
301 TGCCCTACCGCGGGAACCTCTTGCCCTGCTGCTGCTGGGTGCTGGC 350
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 CysSerGluLysAsnProAspCysGlnIleLeuGlyTrpGlyLysMetG1 154
351 GAACGCGAGAATGCCCTACCGTGTGCTGCACTGCGTGAACGTGCGGTGT 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 uAsnGlyGluPheProAspThrIleGlnCysAlaAspValGlnLeuValS 171
401 CTGAGGANGTCTGAGTAAAGTCTATGACCCGCTGTACCACCCACGATG 450
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 erArgGluGluCysGluArgAlaTyrProGlyLysIleThrArgSerMet 187
451 TTCTGCGCCGCGGAGGCGAAGACAGAGAGACTCCTGCAACGCTGACTC 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 ValCysAlaGlyAspLysArgGluGlyAsnAspSerCysGlnGlyAspSe 204
501 TGGGGGCCCCGTGATCTGCAACGGGTACTTGACAGGCGCTTGTCTTTCG 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 rGlyGlyProLeuValCysGlyLysLeuArgGlyIleValSerTrpG 221
551 GAAAGCCCCGTGTGCCCAACTTGCCGTGCCAGGTGTCTACACCAACTC 600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 LysPmetProCysGlySerLysGluLysProGlyValTyrThrAspVal 237
601 TGCAAAATCACTGAGTGATAGAGAAACCGTCCAG 636
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 CysThrHisIleArgTyrPleGlnAsnIleIleArg 249
```

seq_name: sp_vertebrate:093265

seq_documentation_block:

```
ID 093265 PRELIMINARY; PRT; 250 AA.
AC 093265;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE TRYPSINOGEN 1 PRECURSOR (EC 3.4.21.4).
GN TRPL.
OS Pleuronectes americanus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Pleuronectiformes; Pleuronectoidel; Pleuronectidae; Pleuronectes.
RN [1]
```


RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA DOUGLAS S.E., GALLANT J.W.;
RT "Isolation of cDNAs for trypsinogen from the winter flounder,
RT pleuronectes americanus.";
RL J. Mar. Biotechnol. 0:0-0(1998).
DR EMBL; AF012462; AAC32751.1; -.
DR PFAM; PF00089; trypsin; 1.
KW signal; Hydrolase.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 23 250 TRYPSINOGEN 1.
SQ SEQUENCE 250 AA; 27466 MW; D387394D CRC32;

alignment_scores:
Quality: 448.50 Length: 217
Ratio: 2.875 Gaps: 5
Percent Similarity: 71.889 Percent Identity: 42.396

alignment_block:
US-09-030-606-175 x 093265 ..

Align seg 1/1 to: 093265 from: 1 to: 250

1 GCGCAGCCCTGGCAGGGCGGCACTGGTCATGAAAACGAATTGTCCTC 50
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
33 SerArgProTyrMetAlaSerLeuAsnTyrGlyTyrHis...PheCysG1 48
51 GGGCGTCTGGTGCATCCGACAGTGGTGTCTGACGCCGACACTGTTTC 100
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
48 YglYValLeuIleAsnAsnGlnTyrValLeuSerValAlaHisCysTyrP 65
101 AGAAGCTCTACACC.....ATCGGGCTGGGCTGCACAGTCTTGAGGCC 144
||| |||::: |||::: ||| ||| |||::: |||::: |||::: |||:::
65 YrAsnProTyrAlaMetGlnValMetLeuGlyGlnHisAsnLeuArgVal 81
145 GACCAAGCCAGGAGGACGACAGATGTGAGGCCAGCCTTCGCTACGGCA 194
||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
82 ...PheGlnGlyThrGlnGlnLeuMetLysThrAspThrIleIleTyrP 97
195 CCCAGAGTACACAGACTCTTGCTGCCTAACGACCTCATGCTCATCAAGT 244
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
97 sProAsnTyrAspTyrGlnThrLeuAspPheAspIleMetLeuIleLysL 114
245 TGGACGAATCCGTGTCGAGAGTCTGACACCATCCGGAGCATCAGCATGCT 294
||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
114 euTyrHisProValGlnValThrGlnAlaValAlaProIleSerLeuPro 130
295 TCGCAGTGCCTACCGCGGGGAGACTCTTGCTGCTGCTGCTGCTGCTCT 344
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
131 SerSerCysProValGlyGlyThrProCysSerValSerGlyTyrGlyAs 147
345 GCTGGCGAAGGCG.....AGAATGCCCTACCGTGTGCACTGCG 382
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
147 nThrAlaArgAspGlyAspAspValTyrMetProThrLeuLeuGlnCysM 164
383 TGAACGTGTGCGTGTGTCGTGAGANGTCTGCAGTAAGCTCTATGACCCG 432
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
164 etAspValProIleIleAspGlnGlnCysMetLysSerTyrProGly 180
433 CTGTACCAACCCAGCATGTCTGCGCGCGCGGAGGGCAAGACCAAGAGA 482
::: ||| |||::: ||| ||| |||::: ||| ||| |||::: ||| ||| |||:::
181 MetIleSerProArgMetValCysAlaGlyPheMetAspGlySerArgAs 197
483 CTCCTGCAACGGTGAAGTCTGGGGGGGCGCTGATCTGCAACGGGTAAGTGC 532
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
197 pAlaCysAsnGlyAspSerGlySerProLeuValCysArgGlyGlnValT 214
533 AGGGCCTTGTGTTCTTTCGAAAGCCCGGTGTGCGCAACTTGCGTGCCA 582
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
214 hrGlyLeuValSerTyrGlyGlnGly...CysAlaGlnProAsnTyrPro 229
7583 GGTGTCTACACCAACCTCTGCAATTCACCTGAGTGTAGATAGAAAACCGT 632

|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
230 GlyValTyrValLysLeuCysGlnPheHisAlaTyrPleGlnAsnThrLe 246
633 C 633
246 u 246

seq_name: sp_vertibrate:092046

seq_documentation_block:

ID 092046 PRELIMINARY; PRT; 249 AA.
AC 092046; 013261;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE PREPROTRYPSIN PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
OS Dissostichus mawsoni.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Perciformes; Notothenioidel; Nototheniidae; Dissostichus.
[1]

RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE; 97268652.
RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
RT "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in
RT Antarctic notothenioid fish."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3811-3816(1997).
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-I-XAA, LYS-I-XAA.
DR EMBL; U58945; AAB57732.1; -.
DR EMBL; U58835; AAB57728.1; -.
DR PFAM; PF00089; trypsin; 1.
KW signal; Hydrolase.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 249 PROTRYPSIN.
SQ SEQUENCE 249 AA; 27169 MW; 55C59EB4 CRC32;

alignment_scores:
Quality: 447.00 Length: 216
Ratio: 2.847 Gaps: 5
Percent Similarity: 72.685 Percent Identity: 42.130

alignment_block:
US-09-030-606-175 x 092046 ..

Align seg 1/1 to: 092046 from: 1 to: 249

1 GCGCAGCCCTGGCAGGGCGGCACTGGTCATGAAAACGAATTGTCGCTC 50
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
33 SerArgProTyrMetAlaSerLeuAsnTyrGlyTyrHis...PheCysG1 48
51 GGGCGTCTGGTGCATCCGACAGTGGTGTCTGACGCCGACACTGTTTC 100
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
48 YglYValLeuIleAsnAsnGlnTyrValLeuSerValAlaHisCysTyrP 65
101 AGAAGCTCTACACC.....ATCGGGCTGGGCTGCACAGTCTTGAGGCC 144
||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
65 YrAsnProTyrSerMetGlnValIleLeuGlyAspHisAsnLeuArgVal 81
145 GACCAAGCCAGGAGGACGACAGATGTGAGGCCAGCCTTCGCTACGGCA 194
||| |||::: |||::: |||::: |||::: |||::: |||::: |||:::
82 ...PheGlnGlyThrGlnGlnLeuMetLysThrAsnThrIleIleTyrP 97
195 CCCAGAGTACACAGACTCTTGCTGCTAACGACCTCATGCTCATCAAGT 244
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
97 sProSerTyrAspTyrGlnThrLeuAspPheAspIleMetLeuIleLysL 114
245 TGGACGAATCCGTGTGCGAGTCTGACACCATCCGGAGCATCAGCATGCT 294


```
114 eUTYrHisProValGIuValThrGIuAlaValAlaProIleProLeuPro 130
295 TCGCAGTGGCCCTACCGCGGGGAACTCTGCTCTGCTGGCTGGGGCTC 344
131 ThrSerCysProTyrGIyGIyLeuSerCysSerValSerGIyTyrGIyI 147
345 GCTGGCGAAGGC.....AGATGCCTACCGTGTGCTGCACTGCGTGA 385
147 eAlAlYslEuGIyGIyGIuAlaTyrMetProThrLeuGIuGIuGIu 164
386 ACGTGTGGGTGTCTGAGGANGTCTGCAGTAACTCTATGACCCGCTG 435
164 snValProIleValAspGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 180
436 TACCACCCAGCATGTCTGCGCCGCGGAGGCAAGACCAAGACACTC 485
181 IleSerThrThrMetValCysAlaGIyTyrMetGIuGIyGIyAspAl 197
486 CTGCACCGGTGACTCTGGGGGCGCCCTGATCTGCACGCGGTACTTCAG 535
197 aCysAsnGIyAspSerGIySerProLeuValCysAspGIyGIuValGIu 214
536 GCCTGTGTCTTTGGAAGCCCGGTGTGGCCAACCTGGCGTGCAGGT 585
214 YlEuValSerTyrGIyGIuGIy...CysAlaGIuProAsnTyrProGI 229
586 GTCTACACCACTCTGCAAATTCATGAGTGAAGAGAAACCGTC 633
230 ValTyrValYslEuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 245
seq_name: sp_rodent:Q61855
```

```
seq_documentation_block:
ID Q61855 PRELIMINARY; PRT; 261 AA.
AC Q61855;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE TISSUE KALLIKREIN.
GN MGK-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA TADA M., PETERS J., TAKAHASHI S., INOUE H., MIYAKE Y.;
RT "Identification of a tissue kallikrein gene, MGK-6, expressed in a
RL mouse neuroendocrine cell line.";
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; D10464; BAA01257.1;
DR PFM; PF00089; trypsin; 1.
SQ SEQUENCE 261 AA; 28775 MW; 509D1D00 CRC32;
```

```
alignment_scores:
Quality: 447.00 Length: 224
Ratio: 2.829 Gaps: 3
Percent Similarity: 70.536 Percent Identity: 37.946
alignment_block:
us-09-030-606-175 x Q61855 ..
```

```
Align seg 1/1 to: Q61855 from: 1 to: 261
1 GCGCAGCCCTGCGCAGCGCGCAGCTGTCATGAAAAGCAATGTCTGCTC 50
35 SerGIuProTyrGIuValAlaValTyrArgPheThrLysTyrGIuGIu 51
51 GGGCGTCTGTGTCATCCGCGAGTGGGTGCTGTACAGCCGACACTGTTCC 100
51 YGIyIleLeuLeuAsnAlaAsnTyrValLeuThrAlaAlaHisCysHis 68
```

```
101 AGAAGCTCCTACACCATCGGGGTGGCCCTGCACAGTCTTGAGCCGACCA 150
68 snAspLysTyrGIuValTyrLeuGIyLysAsnAsnPheLeuGIuAsp... 83
151 GAGCCAGGAGCCAGATGGTGGAGCCAGCCCTCTCCCTACGGCACCAGAG 200
84 GluProSerAlaGIuHisArgLeuValSerLysAlaIleProHisProAs 100
201 GTACACAGACTCTTGCTC..... 219
100 PheAsnMetSerLeuLeuAsnGIuHisThrProGIuProGIuAspAspT 117
220 ..GCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAGTCT 267
117 YrSerAsnAspLeuMetLeuLeuArgLeuLysLysProAlaAspIleThr 133
268 GACACACATCCGGAGCATCAGCATGCTTCGCACTGCCCTACCGCGGAA 317
134 AspValValLysProIleAspLeuProThrGIuGIuProLysLeuGIySe 150
318 CTCTTGCTCTGCTNTCTGGCTGGGCTCTGCTG.....GCCAAGCGCAGAA 361
150 rThrCysLeuAlaSerGIyTyrGIySerIleThrProValLysTyrGIuT 167
362 TGCTTACCGTGTGCTGACCTGCTGAACGTGTGGGTGCTGAGGANGTC 411
167 YrProAspGIuLeuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 183
412 TGCAGTAGCTCTATGACCCCGGTGTACACCCAGCATGTTCTGCGCCGG 461
184 CysAlaLeuYslAlaHisIleGIuLysValThrAspAspMetLeuGIuGI 200
462 CGAGAGGCAAGACCAAGAGAGACTCCCTGCACCGGTGCTGGGGGCCCC 511
200 YAspMetAspGIyGIyLysAspThrCysAlaGIyAspSerGIyGIyProL 217
512 TGATCTGCACGCGGTACTTGCAGGGCCCTGTGTCTTTGGAAGACCCCG 561
217 euIleCysAspGIyValLeuGIuGIyIleThrSerTyrGIyProSerPro 233
562 TGTGGCCCACTTGCGGTGCGAGGTGTCTACACCAACCTCTGCAATTCAC 611
234 CysGIyLysProAsnValProGIyIleTyrThrArgValLeuAsnPheAs 250
612 TGAGTGGATAGAGAAACCGTC 633
250 nThrTyrIleArgGIuThrMet 257
```

```
seq_name: sp_rodent:O88301
seq_documentation_block:
ID O88301 PRELIMINARY; PRT; 246 AA.
AC O88301;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE SERINE PROTEASE (BSP).
GN MBSP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=skin;
RA MATSUI H., TAKAHASHI T.;
RT "Mouse serine protease preferentially expressed in brain.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=skin;
RA MEIER N., DEAR T.N., BOEHM T.;
RT "A novel serine protease expressed in the hair follicle.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
```


[illegible]

```
alignment_scores:
    Quality: 431.50      Length: 211
    Ratio: 2.839      Gaps: 4
    Percent Similarity: 72.038      Percent Identity: 39.336

alignment_block:
```

```

US-09-030-606-175 x Q9Z1R9      ..
Align seg 1/1 to: Q9Z1R9 from: 1 to: 246

```

| | | |
|-----|--|-----|
| 7 | CCCTGGCAGGCGGCACCTGGTCATGGAAGAAACGAATTTGTTCTGCTGGGGCGT | 56 |
| 36 | ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlyse | 51 |
| 57 | CCGTGGTCATCCGCAGTGGGTGGCTGTACGCCGACACACTGTTTCCAGAACT | 106 |
| 51 | rLeuIleAsnAspGlnTrpValValSerAlaAlaHisCysTyrLysThra | 68 |
| 107 | CCATACCATCCGGGCTGGGCTGGCCAGTCTTGAGGCCGACCAAGACCA | 156 |
| 68 | rGlnGlnValArgLeuGlyGlnHisAsnIleAsnVal...LeuGlnGly | 83 |
| 157 | GGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGACTACAA | 206 |
| 84 | AsnGlnGlnPheIleAsnAspAlaAlaLysIleIleLysHisProAsnPheAs | 100 |
| 207 | CAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG | 256 |
| 100 | naArgLysThrLeuAsnAsnAspIleMetLeuIleLysLeuSerSerProv | 117 |
| 257 | TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCCAGTGCCT | 306 |
| 117 | alThrLeuAsnAlaArgValAlaThrValAlaLeuProSerSerCysAla | 133 |
| 307 | ACCGGGGGGAACCTTGCCCTCGTNTCTGGCTGGGT....CTGCTGGC | 350 |
| 134 | ProAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLeuSerPh | 150 |
| 351 | GAACGGCAGAATGCCCTACCCTGCTGCACATGCCGTGAACGTGCGTGCTGT | 400 |
| 150 | egLysValSerGluProAspLeuLeuGlnCysLeuAspAlaProLeuLeuP | 167 |
| 401 | CTGAGGANGTCTGCAGTAAGCTTATGACCCCGCTGTACCACCCAGCATG | 450 |
| 167 | roGlnAlaAspCysGlnAlaSerTyrProGlyLysIleThrGlyAsnMet | 183 |
| 451 | TTCTGGCCCGGGCGAGGCAAGACCAAGAGACTCCTGCAACGCTGACTC | 500 |
| 184 | ValCysAlaGlyPheLeuGlnGlyGlyLysAspSerCysGlnGlyAspse | 200 |
| 501 | TGGGGGGCCCTGATCTGCAACGGGTACTTGACAGGCGCTGTGTCTTCG | 550 |
| 200 | rgLysGlyProValValCysAsnGlyIleLeuGlnGlyIleValSerTrpG | 217 |
| 551 | GAAAGCCCCGTGTGGCCAACTTGGCGGTGCCAGGTGTTACACCAACCTC | 600 |
| 217 | LYTyrGly...CysAlaLeuProAspAsnProGlyValTyrThrLysVal | 232 |
| 601 | TGCAAAATTCACCTGAGTGATAGAGAAAACCGTC | 633 |
| 233 | CysAsnTyrValAspTrpIleGlnAspThrIle | 243 |

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:43 ; Search time 1809.22 Seconds
(without alignments)
1272.342 Million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167
Sequence: 1 GCGCAGCCCTGCGCAGCGGC.....NTAAGAGAAGCGCAAAAAA 1167

Scoring table: IDENTITY_NDC
Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 555.6 | 47.6 | 777 | 48 | AI557281 | AI557281 PT2.1_15_ |
| 2 | 403.2 | 34.6 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 3 | 393 | 33.7 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 4 | 348 | 29.8 | 404 | 34 | AA503963 | AA503963 nh39a01.s |
| 5 | 286.6 | 24.6 | 576 | 34 | AA533140 | AA533140 nj46h06.s |
| 6 | 221 | 18.9 | 722 | 48 | AI557025 | AI557025 PT2.1_10_ |
| 7 | 217.2 | 18.6 | 259 | 35 | AA565161 | AA565161 nk52c07.s |
| 8 | 210.8 | 18.1 | 229 | 35 | AA552779 | AA552779 nk57a10.s |
| 9 | 203 | 17.4 | 218 | 50 | AI674521 | AI674521 wc39g02.x |
| 10 | 201.2 | 17.2 | 241 | 36 | AA603977 | AA603977 nc47a09.s |
| 11 | 174 | 14.9 | 188 | 50 | AI675815 | AI675815 wb97b12.x |
| 12 | 164 | 14.1 | 216 | 32 | AA336074 | AA336074 EST40886 |
| 13 | 161.8 | 13.9 | 191 | 35 | AA595489 | AA595489 nc034h04.s |
| 14 | 132 | 11.3 | 457 | 45 | AI385433 | AI385433 m185b05.y |
| 15 | 118.2 | 10.1 | 498 | 26 | W73168 | W73168 zd55e11.r1 |
| 16 | 116 | 9.9 | 505 | 45 | AI391329 | AI391329 mb71h07.y |
| 17 | 114 | 9.8 | 496 | 39 | AA864127 | AA864127 vn42g07.r |
| 18 | 112.6 | 9.6 | 507 | 35 | AA583052 | AA583052 nr80g09.s |
| 19 | 106.4 | 9.1 | 517 | 31 | AA293027 | AA293027 zt54a12.r |
| 20 | 100.2 | 8.6 | 590 | 33 | AA411252 | AA411252 zt33b03.r |
| 21 | 99.2 | 8.5 | 759 | 45 | AI326340 | AI326340 m161e08.x |
| 22 | 99 | 8.5 | 321 | 31 | AA302930 | AA302930 EST113097 |
| 23 | 98.4 | 8.4 | 686 | 46 | AI415008 | AI415008 mb71h07.x |
| 24 | 98 | 8.4 | 586 | 26 | W73140 | W73140 zd55e11.s1 |
| 25 | 97.4 | 8.3 | 626 | 36 | AA643312 | AA643312 nr59b08.s |
| 26 | 96 | 8.2 | 333 | 49 | AI636241 | AI636241 tz93c03.x |
| 27 | 92.6 | 7.9 | 356 | 39 | AA838788 | AA838788 ai87g08.s |
| 28 | 91.4 | 7.8 | 212 | 28 | C16642 | C16642 C16642 Clon |
| 29 | 90.2 | 7.7 | 420 | 43 | AI226226 | AI226226 ue88f06.y |
| 30 | 90 | 7.7 | 585 | 35 | C23111 | C23111 C23111 Japa |
| 31 | 90 | 7.7 | 504 | 49 | AI620091 | AI620091 ty48d10.x |
| 32 | 89.8 | 7.7 | 480 | 39 | AA846771 | AA846771 aj41f01.s |
| 33 | 89.8 | 7.7 | 367 | 41 | AI002101 | AI002101 ot38d03.s |
| 34 | 89.6 | 7.7 | 234 | 33 | AA452459 | AA452459 zx29g09.r |
| 35 | 89 | 7.6 | 400 | 31 | AA293231 | AA293231 zt26g09.r |
| 36 | 89 | 7.6 | 539 | 34 | AA477689 | AA477689 zu44a12.r |
| 37 | 87.8 | 7.5 | 467 | 28 | AA073833 | AA073833 mj99h09.r |
| 38 | 87.8 | 7.5 | 460 | 38 | AA791893 | AA791893 vs54h07.r |
| 39 | 87.4 | 7.5 | 519 | 33 | AA401397 | AA401397 zu68b01.s |
| 40 | 87.4 | 7.5 | 583 | 40 | AA921373 | AA921373 ak56d12.s |
| 41 | 87.4 | 7.5 | 493 | 45 | AI324874 | AI324874 m185b05.x |
| 42 | 87.2 | 7.5 | 541 | 39 | AA844955 | AA844955 ak61d09.s |
| 43 | 87.2 | 7.5 | 324 | 43 | AI177474 | AI177474 EST221106 |
| 44 | 87.2 | 7.5 | 599 | 43 | AI237604 | AI237604 EST234166 |
| 45 | 87.2 | 7.5 | 457 | 46 | AA998593 | AA998593 UI-R-CO-1 |

ALIGNMENTS

RESULT 1
AI557281
LOCUS AI557281 777 bp mRNA
DEFINITION PT2.1_15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557281
NID g4489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

| | |
|-----------|--|
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 777) |
| AUTHORS | Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J. and Hood, L. |
| TITLE | Prostate Cancer Expression Profiling by cDNA Sequencing Analysis |
| JOURNAL | Unpublished (1999) |
| COMMENT | On May 18, 1998 this sequence version replaced gi:3138767. |

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .777 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
CDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."

```

| | | | | |
|---------------------------|--------|---------------------|------------|-------------|
| Query Match | 47.68; | Score 555.6; | DB 48; | Length 777; |
| Best Local Similarity | 91.88; | Pred. No. 1.3e-128; | | |
| Matches 637; Conservative | 0; | Mismatches 46; | Indels 11; | Gaps 5; |

| | | | | |
|----|-----|--|---------------------------|----|
| QY | 2 | CGCAGCCCTGGCAGCGCGGCACCTGGTGCATGGAAAA | CGAATTGTTCTGCTGGGGCGTCTGG | 61 |
| | | | | |
| Db | 21 | CGCAGCCCTGGCAGCGCGGCACCTGGTGCATGGAAAA | CGAATTGTTCTGCTGGGGCGTCTGG | 80 |
| | | | | |
| QY | 62 | TGCATCCGAGTGGGTGCTGTGCAGCCGCACACTGTTTCCAGAACTCCTACACCATCGGGC | 121 | |
| | | | | |
| Db | 81 | TGCATCCGAGTGGGTGCTGTGCAGCCGCACACTGTTTCCAGAACTCCTACACCATCGGGC | 140 | |
| | | | | |
| QY | 122 | TGGGCGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGGAGCCAGATGGTGGAGGCCACCC | 181 | |
| | | | | |
| Db | 141 | TGGGCGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGGAGCCAGATGGTGGAGGCCACCC | 200 | |
| | | | | |
| QY | 182 | TCTCCGTACGGGCACCCAGAGTACACACAGACTTGTCTCGCTAACGACCTCATGCTCATCA | 241 | |
| | | | | |
| Db | 201 | TCTCCGTACGGGCACCCAGAGTACACACAGACCTTGTCTCGCTAACGACCTCATGCTCATCA | 260 | |
| | | | | |
| QY | 242 | AGTTGGACGAATCCGTGTGCCAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCGAGT | 301 | |
| | | | | |
| Db | 261 | AGTTGGACGAATCCGTGTGCCAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCGAGT | 320 | |
| | | | | |
| QY | 302 | GCCCTACCGCGGGGGAACCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 361 | |
| | | | | |
| Db | 321 | GCCCTACCGCGGGGGAACCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 380 | |
| | | | | |
| QY | 362 | TGCCTACCGGTGCTGCACTGCGGTGAACGTGTGGGTGTCTGAGGANGTCTGCAGTAAGC | 421 | |
| | | | | |
| Db | 381 | TG-CTACCGTGTGCACTGCGGTGAACGTGTGGGTGTCTGAGGANGTCTGCAGTAAGC | 439 | |
| | | | | |
| QY | 422 | TCTATGACCCGCTGTACCAACCCAGACATGTTCTGCGCCGGCGGAGGGCAAGACCAAGG | 481 | |
| | | | | |
| Db | 440 | TCTATGACCCGCTGTACCAACCCAGACATGTTCTGCGCCGGCGGAGGGCAAGACCAAGG | 499 | |
| | | | | |
| QY | 482 | ACTCCTGCACAGGTGACTCTGGGGGGCCCTGATCTGCACACGGGTACTTGACAGGGCCTTG | 541 | |
| | | | | |
| Db | 500 | ACTCCTGCACAGGTGACTCTGGGGGG-CCCTGATNTGNAACGGGGNCTTGACAGGGCCTTG | 558 | |

QY 542 TGCTTTTGGAAAAGCCCCGTGTGGCCACTTGTGGCGTGCAGTGTCTACACCAACTCT 601
||||| ||||| ||||| | ||| ||||| ||||| | ||| | ||
Db 559 TGCTTTTGGAAAAGCCCCGTGTGG--CAAGTTGGCGTGCAGTGTCTACACAAC-----CT 611
QY 602 GCAAAATTCACGTAGTGGATAGAGAAAACCGTCCAGNCCAGTTACTCTGGGACTGGGA 661
||| | ||| ||||| ||||| ||||| | ||||| ||||| ||| |
Db 612 CTGAATTAATGAGTGGATAGANAAAACGCTCAG--CAGTTACTNTGGGNACTGGNA 669
QY 662 CCCATGAATTGACCCCCCAATATACATCTGCCGA 695
||||| ||||| ||||| ||||| ||| | | |
Db 670 NCCATTAATTGNCCTCCCAAAATANATCTGGGGAA 703

| | |
|------------|---|
| RESULT | 2 |
| AI686689/c | |
| LOCUS | AI686689 |
| DEFINITION | tus3g11.x1 NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2253092 3 similar to TR:Q92046 Q92046 PREPROTRYPsin PRECURSOR ; mRNA sequence. |
| ACCESSION | AI686689 |
| NID | 94897983 |
| VERSION | AI686689.1 |
| KEYWORDS | EST. |
| SOURCE | human. |

| | |
|-----------|--|
| ORGANISM | Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominini; Homo. |
| REFERENCE | 1 (bases 1 to 569) |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | On Jun 5, 1998 this sequence version replaced gi:3189584. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from GIBCO
High quality sequence stop: 444..

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

```

| | | | | |
|---------------------------|-------|--------|--------------------|-------------|
| BASE COUNT | 125 a | 150 c | 170 g | 124 t |
| ORIGIN | | | | |
| Query Match | | 34.6%; | Score 403.2; | DB 50; |
| Best Local Similarity | | 98.8%; | Pred. No. 1.1e-90; | Length 569; |
| Matches 405; Conservative | | 0; | Mismatches 5; | Indels 0; |
| | | | | Gaps 0; |

QY 151 GAGCCAGGAGCCAGATGTTGGAGGCCAGCCCTCTCCGTACGGCACCACAGTACACAGA 210
|||||
Db 569 GAGCCAGGAGGCCAGATGTTGGAGGCCAGCCCTCTCCGTACGGCACCACAGTACACAGA 510
QY 211 CTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGAC 270
|||||
Db 509 CCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGAC 450
QY 271 ACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCTTACCAGCGGGGAACCTTTCCTGCTN 330
|||||
Db 449 ACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCTTACCAGCGGGGAACCTTTCCTGCTN 390
QY 331 TCTGGCTGGGCTGCTGGCGAAGCAGCAGAAATGCTACCGTGTGCACTGCGTGAACGTG 390
|||||
Db 389 TCTGGCTGGGCTGCTGGCGAAGCAGCAGAAATGCTACCGTGTGCACTGCGTGAACGTG 330
QY 391 TCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATG 450
|||||
Db 329 TCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATG 270
QY 451 TTCTGCGCGCGGAGGCGAAGACCAAGAGACTCTGCAACGGGTGACTGTGGGGGCC 510
|||||
Db 269 TTTGCGCGCGGAGGCGAAGACCAAGAGACTCTGCAACGGGTGACTGTGGGGGCC 210
QY 511 CTGATCTGCAACGGGTGACTTGACGGGCTTGTCTTTCGAAAAAGCCCC 560
|||||
Db 209 CTGATCTGCAACGGGTGACTTGACGGGCTTGTCTTTCGAAAAAGCCCC 160

RESULT 3
AA551449 415 bp mRNA EST 05-SEP-1997
LOCUS nJ55e05.s1 NCI_CGAP_P19 Homo sapiens cDNA clone IMAGE:996416
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html

FEATURES
Source
Insert Length: 640 Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.
Location/Qualifiers
1. 415

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_id="NCI_CGAP_P19"
/sex="male"

/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."

BASE COUNT 78 a 125 c 129 g 83 t
ORIGIN
Query Match 33.7%; Score 393; DB 35; Length 415;
Best Local Similarity 98.1%; Pred. No. 3.3e-88;
Matches 407; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 137 TTGAGGCCAGCAGAGCCAGGAGCCAGATGTTGAGGCCAGCCCTCTCCGTACGGCACC 196
|||||
Db 2 TCGACGCCAGCAGAGCCAGGAGCCAGCAGCGGTGAGGCCAGCCCTCTCCGTACGGCACC 61
QY 197 CAGAGTACAACAGACTCTGCTCGCTTAACGACCTCATGCTCAAGTTGGACGAATCCG 256
|||||
Db 62 CAGAGTACAACAGACTCTGCTCGCTTAACGACCTCATGCTCAAGTTGGACGAATCCG 121
QY 257 TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCCAGTGGCCCTACCGGGGA 316
|||||
Db 122 TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCCAGTGGCCCTACCGGGGA 181
QY 317 ACTCTTGCCCTCTGCTTGTGGCTGGGTGCTGCTGGCGAACGGCAGAAATGCCCTACCGTGTGC 376
|||||
Db 182 ACTCTTGCCCTCTGCTTGTGGCTGGGTGCTGCTGGCGAACGGCAGAAATG-CTACCGTGTGC 240

QY 377 ACTGCGTGAACGTGTCGGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGT 436
|||||
Db 241 AGTGGTGAACGTGTCGGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGT 300
QY 437 ACCACCCAGCATGTTCTGCGCGCGGAGGCGAAGACCAAGAGACTCTGCAACGGTG 496
|||||
Db 301 ACCACCCAGCATGTTCTGCGCGCGGAGGCGAAGACCAAGAGACTCTGCAACGGTG 360

QY 497 ACTCTGGGGGCCCTGATCTGCAACGGGTACTTGACGGCCCTTGTCTTCCG 551
|||||
Db 361 ACTCTGGGGGCCCTGATCTGCAACGGGTACTTGACGGCCCTTGTCTTCCG 415

RESULT 4
AA503963 404 bp mRNA EST 20-AUG-1997
LOCUS nh39a01.s1 NCI_CGAP_P15 Homo sapiens cDNA clone IMAGE:954696
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; contains MSRI.b2 MSRI repetitive element ; mRNA sequence.
ACCESSION AA503963
NID 92238930
VERSION AA503963.1 GI:2238930
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 8, 1995 this sequence version replaced gi:801235.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert length: 1244 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 144.

FEATURES

source

Location/Qualifiers
1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954696"
/clone_1lb="NCI_CGAP_Pr5"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from normal prostatic epithelial cells, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."
BASE COUNT 64 a 91 c 159 g 90 t
ORIGIN

Query Match 29.8%; Score 348; DB 34; Length 404;
Best Local Similarity 93.5%; Pred. No. 5.3e-77;
Matches 371; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

OY 502 GGGGGCCCCCTGATCTGCACGGGTAATTCACAGGGCCCTGTGCTTTCGGAAAAAGCCCCG 561
||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 404 GGGGGCCCCCTGAATCTGCACGGGTAATTCACAGGGCCCTGTGCTTTCGGAAAAAGCACCC 345
OY 562 TGTGGCCCACTTGGCGTGCAGGTTCTACACCAACCTCTGCAATTCA-CTGAGTGGAT 620
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 344 TGTGGCCCAAGTTGGCGTGCAGGTTCTACACCAACCTCTGCAAAATTCACCTGAGTGGAT 285
OY 621 AGAGAAAACCGTCCAGNCCAGTTAATCTGGGACTGGGAACCCATGAATTGACCCCCA 680
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 284 AGAGAAAACCGTCCAGGCAAGTACTGTGGGACTGGGAACCCATGAATTGACCCCCA 225
OY 681 AATACATCTGCGGAANGAATTCAGGAATATCTGTCCAGCCCCCTCCTCCTCAGGCC 740
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 224 AATACATCTGCGGAANGAATTCAGGAATAGGTGTCCAGCCCCCTCCTCCTCAGGCC 165
OY 741 AGAGTCCAGGCCCCCAGCCCCCTCCTCCTCAAACCAAGGTAACAGATCCCCAGCCCCCTC 800
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 164 AGAGTCCAGGCCCCCAGCCCCCTCCTCCTCAAACCAAGGTAACAGAACCCAGCCCCCTC 105
OY 801 CTCCCTCAGACCCAGGAGTCCAGAGCCCCCAGCCCCCTCCTCCTCAGACCCAGAGTCCA 860
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 104 CTCCCTCAGACCCAGGAGTCCAGAGCCCCCAGCCCCCTCCTCCTCAGACCCAGAGTCCA 45
OY 861 GCCCCTCCTCCTCAGACGCGAGGATCCAGACCCCC 897
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 44 GCCCCTCCTCCTCAGACCCAGGAGTCCAGACCCCC 8

RESULT 5

AA533140 576 bp mRNA EST 21-AUG-1997
LOCUS nJ46h06.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:995579
DEFINITION similar to TR:G198491 G198491 KALIKREIN ; contains MSRL.tl MSRL
repetitive element ;, mRNA sequence.

ACCESSION

AA533140

NID 92277236

VERSION

AA533140.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 576)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

On Sep 12, 1996 this sequence version replaced gl:1406821.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquiqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert length: 885 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 256.

FEATURES

source

Location/Qualifiers
1. 576
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:995579"
/clone_1lb="NCI_CGAP_Pr9"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMPI0; mRNA made from normal prostatic epithelial cells, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."

BASE COUNT 169 a 156 c 139 g 102 t 10 others
ORIGIN

Query Match 24.6%; Score 286.6; DB 34; Length 576;
Best Local Similarity 94.7%; Pred. No. 1.2e-61;
Matches 306; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

OY 472 GACCAGAAGACTCTCTGCAACGGTGACT-CTGGGGGCCCCCTGATCTGCAACGGGTACTT 530
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 42 GACCAGAAGACTCTCTGCAACGGTGACTCTGTGGGGCCCCCTGATCTGCAACGGGTACTT 101
OY 531 GCAGGGCCTTGTGCTTTTGGAAAAAGCCCCGTGGCCAACTTGGCGTGCAGGTGCTTA 590
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 102 GCAGGGCCTTGTGCTTTTGGAAAAAGCCCCGTGGCCAACTTGGCGTGCAGGTGCTTA 161
OY 591 CACCAACTCTGCATAATTCAGTGAATAGAGAAAAACCGTCCAGNCCAGTTAACTCTG 650
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 162 CACCAACTCTGCATAATTCAGTGAATAGAGAAAAACCGTCCAGNCCAGTTAACTCTG 221
OY 651 GGGACTGGGAACCATGAATTGACCCCCCAATATCATCTGCGGAANGAATTGAGGATA 710
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 222 GGGACTGGGAACCATGAATTGACCCCCCAATATCATCTGCGGAANGAATTGAGGATA 281
OY 711 TCTGTCCAGACCCCTCCTCCTCAGGCCCAAGAGTCCAGGCCCAAGCCCTCCTCCT 770
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 282 TCTGTCCAGACCCCTCCTCCTCAGGCCCAAGAGTCCAGATCCCCAGAACTACTTCT 341
OY 771 CAACCAAGGGTACAGATCCCA 793
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 342 TCAACCAAGGGTACAGATTCCA 364

RESULT 6

AI557025

LOCUS

AI557025

DEFINITION

PT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.

ACCESSION

AI557025

NID

94489388

VERSION AI557025.1 GI:4489388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 722)
AUTHORS Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.
Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
location/Qualifiers
1. 722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 146 a 156 c 195 g 176 t 49 others
ORIGIN
Query Match 18.9%; Score 221; DB 48; Length 722;
Best Local Similarity 94.8%; Pred. No. 2.9e-45;
Matches 235; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
QY 103 AACTCCTACACATCGGGCTGGGCTGCACAGTCTTGAGGCGCAGACCAAGCCAGGAGC 162
Db 115 AGCTTCTACACATCGGGCTGGGCTGCACAGTCTTGAGGCGCAGACCAAGCCAGGAGC 174
QY 163 CAGATGGTGAGGCGCAGCTCTCCGTACGCGACCCAGAGTACACAGACTCTGCTCGCT 222
Db 175 CAGATGGTGAGGCGCAGCTCTCCGTACGCGACCCAGAGTACACAGACTCTGCTCGCT 234
QY 223 AAGCAGCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGC 282
Db 235 AAGCAGCTCATGCTCATNAAGTTGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGC 294
QY 283 ATCAGCATGCTTCGACAGTGCCCTACCGCGGGAAGTCTGCC-TCGTTCTGCTGGG 341
Db 295 ATCAGCATGCTTCGAGNAGTGCCCTACCGGTGGGGAAGTCTTNCCTTCGTTCTGCTGGG 354
QY 342 TCTGCTGG 349
Db 355 TCTGCTGG 362
RESULT 7
AA565161/c 259 bp mRNA EST 08-SEP-1997
LOCUS nk52c07.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017132
DEFINITION similar to contains MSRI.b3 MSRI MSRI repetitive element ;, mRNA sequence.
ACCESSION AA565161
NID 92336800
VERSION AA565161.1 GI:2336800
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Euthera; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 259)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397569.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 458 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 258.
location/Qualifiers
1. 259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q21"
/clone="IMAGE:1017132"
/clone_lib="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMPI0; mRNA made from prostate intraepithelial neoplasia (high-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."
BASE COUNT 56 a 41 c 106 g 56 t
ORIGIN
Query Match 18.6%; Score 217.2; DB 35; Length 259;
Best Local Similarity 91.8%; Pred. No. 1.7e-44;
Matches 225; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 915 ACCCAGGGTGACGCGCCCAACCCCTCCTCCTCAGAGTCAAGTCCAGCCCAAC 974
Db 245 AACCAAGGTACAGATCCCAAGCCCTCCTCCTCAGAGTCAAGTCCAGCCCAAC 186
QY 975 CCCTGTTCCCAAGACCCAGAGTNCAGGTCCAGCCCTCCTCCTCAGACCGCGT 1034
Db 185 CCCTCCTCCCAAGACCCAGAGTNCAGGTCCAGCCCTCCTCCTCAGACCGCGT 126
QY 1035 CCAATGCCACTAGANTNTCCCTGTACACAGTGCCTGTTGGCANGTTGACCAACCT 1094
Db 125 CCAATGCCACTAGACTCTCCCTGTACACATGCCCCCTGTGTGACAGTCAACCT 66
QY 1095 TACCAGTTGGTTTTCATTTTGTGCTTCCCTAGATCCAGAAATAAGTNTAAGAG 1154
Db 65 TACCAGTTGGTTTTCATTTTGTGCTTCCCTAGATCCAGAAATAAGTNTAAGAG 6
QY 1155 AAGCG 1159
Db 5 AAGCG 1
RESULT 8
AA552779/c 229 bp mRNA EST 08-SEP-1997
LOCUS nk57a10.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017594
DEFINITION similar to contains MSRI.b3 MSRI repetitive element ;, mRNA sequence.

ACCESSION AA552779
 NID 92323033
 VERSION AA552779.1 GI:2323033
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominini; Homo.
 1 (bases 1 to 229)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 25, 1995 this sequence version replaced gi:637760.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nh.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 419 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 215.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .229 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1017594"
/clone_lib="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from prostate intraepithelial neoplasia (high-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp"

```

| | | | | | | | | | |
|-----------------------|--------|--------------|----|------------|-----|--------|----|------|----|
| BASE COUNT | 49 | a | 38 | c | 94 | g | 48 | t | |
| ORIGIN | | | | | | | | | |
| Query Match | 18.1%; | | | | | | | | |
| Best Local Similarity | 93.9%; | | | | | | | | |
| Matches | 215; | Conservative | 0; | Mismatches | 14; | Indels | 0; | Gaps | 0; |

| | | | | | | | | |
|----|------|--------------|-----------------|----------|------------------|--------------|----------------|------|
| QY | 931 | CCCCAACCCCTC | NTCCNTGAGT | CAGAGT | CCAGAG | CCCCCAACCCCT | CGTTCCCGAGAC | 990 |
| | | | | | | | | |
| Db | 229 | CCCCAGCCCTC | CTCCTCAGACC | AGAGT | CCAAAGCCCAACCCCT | CTTCCCGAGAC | 170 | |
| QY | 991 | CCAGAGTNCAGT | CCAGCCCTCCTCCT | CAGACC | AGCGTCCATG | CCACCTAGAN | 1050 | |
| | | | | | | | | |
| Db | 169 | CCAGAGTCCAGT | ACCAGCCCTCCTCCT | CAGACC | AGCGGTCCATG | CCACCTAGAC | 110 | |
| QY | 1051 | TNTCCCTGTAC | ACAGTGC | CCCCCTTG | TGCGANGT | TGACCAACCTT | ACCAGTTGGTTTTC | 1110 |
| | | | | | | | | |
| Db | 109 | TCTCCCTGTAC | ACAGTGC | CCCCCTTG | TGCGACGTTG | ACCACAACCTT | ACCAGTTGGTTTTC | 50 |
| QY | 1111 | ATTTTGTGCT | CTTCCCTT | AGATCCAG | AAATAAAGT | NTAAGAGAGCG | 1159 | |
| | | | | | | | | |
| Db | 49 | ATTTTGTGCT | CTTCCCTT | AGATCCAG | AAATAAAGT | CTAAGAGAGCG | 1 | |

| | | | | | |
|------------|------------|---------------|--------------|------------|-----------------|
| RESULT | 9 | | | | |
| AI674521/c | | | | | |
| LOCUS | AI674521 | 218 bp | mRNA | EST | 19-MAY-1999 |
| DEFINITION | WG39g02.x1 | NCI_CGAP_Pr28 | Homo sapiens | CDNA clone | IMAGE:2321042 3 |

| | |
|-----------|--|
| ACCESSION | similar to contains MSRL.t2 MSRL repetitive element ; , mRNA sequence. |
| NID | AI674521 |
| VERSION | g4875001 |
| KEYWORDS | AI674521.1 GI:4875001 |
| SOURCE | EST. |
| ORGANISM | human. Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| AUTHORS | 1 (bases 1 to 218) |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP). |
| JOURNAL | Tumor Gene Index |
| COMMENT | Unpublished (1997) On Mar 16, 1998 this sequence version replaced gi:2961915. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .218 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2321042"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

```

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 17.4%; | Score 203; | DB 50; | Length 218; |
| Best Local Similarity | 94.9%; | Pred. No. 5.6e-41; | | |
| Matches 206; Conservative | 0; | Mismatches 11; | Indels 0; | Gaps 0; |

| | | | |
|----|------|---|------|
| QY | 944 | TCCNTCAGAGTCAGAGGTCACAAAGCCCCCAACCCTTGTTCCCCAGAGACCAGAGGTNCAGG | 1003 |
| | | | |
| Db | 217 | TCCTCAGACTCAGAGGTCACAAGCCCCCAACCCTTCTTCCCAGAGACCAGAGGTCCAGG | 158 |
| QY | 1004 | TCCCAAGCCCCCTCCTCCTCCTCAGAACCCAGCGGTCCAATGCCACCTAGANTNCCCTGTACAC | 1063 |
| | | | |
| Db | 157 | TCCCAAGCCCCCTCCTCCTCCTCAGAACCCAGCGGTCCAATGCCACCTAGACTCTCCCTGTACAC | 98 |
| QY | 1064 | AGTGGCCCCCTTGTCGCANHTGAACCCAACCTTACCAAGTTGGTTTTTCATTTTTTGTCCCT | 1123 |
| | | | |
| Db | 97 | AGTGGCCCCCTTGTCGCACGTTGACCCCAACCTTACCAGTTGGTTTTTCATTTTTTGTCCCT | 38 |
| QY | 1124 | TTCCCTTAGATCCAGAATAAAGTNTAAGAGAAGCGC | 1160 |
| | | | |
| Db | 37 | TTCCCAAAGATCCAGAATAAAGTCTAAGAAAAGCGC 1 | |

QY 1124 TTCCCTAGATCCAGAAATAAGTNTAAGAGAGCCG 1160
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 37 TTCCCAAGATCCAGAAATAAGTCTAAGAAAGCCG 1

[illegible]

| | | | | | | |
|---------------------------|------|--|--------|------|-----|-------------|
| RESULT | 11 | AI675815 | 188 bp | mRNA | EST | 19-MAY-1999 |
| LOCUS | | AI675815/c | | | | |
| DEFINITION | | wb97b12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2313599 3', | | | | |
| ACCESSION | | AI675815 | | | | |
| NID | | 94876295 | | | | |
| VERSION | | AI675815.1 | | | | |
| KEYWORDS | | EST. | | | | |
| SOURCE | | human. | | | | |
| ORGANISM | | Homo sapiens | | | | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | | |
| AUTHORS | | Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | |
| TITLE | | 1 (bases 1 to 188) | | | | |
| JOURNAL | | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . | | | | |
| COMMENT | | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), | | | | |
| | | Tumor Gene Index | | | | |
| | | Unpublished (1997) | | | | |
| | | On Dec 20, 1995 this sequence version replaced gi:1133540. | | | | |
| | | Contact: Robert Strausberg, Ph.D. | | | | |
| | | Tel: (301) 496-1550 | | | | |
| | | Email: Robert_Strausberg@nih.gov | | | | |
| | | Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. | | | | |
| | | Emmert-Buck, M.D., Ph.D. | | | | |
| | | CDNA Library Preparation: M. Bento Soares, Ph.D. | | | | |
| | | CDNA Library Arrayed by: Greg Lennon, Ph.D. | | | | |
| | | DNA Sequencing by: Washington University Genome Sequencing Center | | | | |
| | | Clone distribution: NCI-CGAP clone distribution information can be | | | | |
| | | found through the I.M.A.G.E. Consortium/BLN at: | | | | |
| | | www.bio.lnl.gov/db/rp/lmage/lmage.html | | | | |
| FEATURES | | Seq primer: -40UP from Gibco. | | | | |
| source | | Location/Qualifiers | | | | |
| | | 1.188 | | | | |
| | | /organism="Homo sapiens" | | | | |
| | | /db_xref="taxon:9606" | | | | |
| | | /clone="IMAGE:2313599" | | | | |
| | | /clone_lib="NCI_CGAP_Pr28" | | | | |
| | | /sex="male" | | | | |
| | | /dev_stage="adult" | | | | |
| | | /lab_host="DH10B" | | | | |
| | | /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) | | | | |
| | | with a modified polylinker; Plasmid DNA from the | | | | |
| | | normalized library NCI_CGAP_Pr22 was prepared, and ss | | | | |
| | | circles were made in vitro. Following HAP purification, | | | | |
| | | this DNA was used as tracer in a subtractive hybridization | | | | |
| | | reaction. The driver was PCR-amplified cDNAs from a pool | | | | |
| | | of 5,000 clones made from the same library (cloneIDs | | | | |
| | | 985608-986759, 1101192-1101959, and 1217928-1220615). | | | | |
| | | Subtraction by Bento Soares and M. Fatima Bonaldo." | | | | |
| BASE COUNT | | 43 a 31 c 61 g 53 t | | | | |
| ORIGIN | | | | | | |
| Query Match | | 14.9%; Score 174; DB 50; Length 188; | | | | |
| Best Local Similarity | | 97.28; Pred. No. 8.9e-34; | | | | |
| Matches 174; Conservative | | 0; Mismatches 5; Indels 0; Gaps 0; | | | | |
| OY | 989 | ACCAGAGTNCAGTCCAGCCCTCCTCCCTCAGACCCAGCGGTCCAAATGCCACTAG | 1048 | | | |
| Db | 188 | ACCCAGAGGTCCAGTCCAGCCCTCCTCCCTCAGACCCAGCGGTCCAAATGCCACTAG | 129 | | | |
| OY | 1049 | ANTNTCCCTGTACAGAGTGCCTGTGTGGCAGTGTGACCCCAACCTTACCAAGTGT | 1108 | | | |
| Db | 128 | ACTCTCCCTGTACAGAGTGCCTGTGTGGCAGTGTGACCCCAACCTTACCAAGTGT | 69 | | | |
| OY | 1109 | TCATTTTGTGTCCTTTCCCTTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA | 1167 | | | |
| Db | 68 | TCATTTTGTGTCCTTTCCCTTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA | 10 | | | |


```
RESULT 12
AA336074      216 bp      mRNA      EST      21-APR-1997
LOCUS
DEFINITION   AA336074      Endometrial tumor Homo sapiens cDNA 5' end similar to
similar to kallikrein family, mRNA sequence.
ACCESSION
NID          AA336074
VERSION      AA336074.1  GI:1988560
KEYWORDS
SOURCE
ORGANISM     Homo sapiens
              human.
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 216)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fletschmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudex,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrle,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL      Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE
COMMENT      On Apr 14, 1993 this sequence version replaced gi:692773.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES
    source
        1..216
        /organism="Homo sapiens"
        /db_xref="ATCC (Inhost):138014"
        /db_xref="taxon:9606"
        /clone_lib="Endometrial tumor"
        /sex="female"
        /dev_stage="adult"
        /note="Organ: endometrium; Vector: pBluescript SK-;
        Site_1: EcoRI; Site_2: XhoI"
BASE COUNT   45 a      65 c      63 g      38 t      5 others
ORIGIN
Query Match      14.1%; Score 164; DB 32; Length 216;
Best Local Similarity 97.0%; Pred. No. 2.9e-31;
Matches 164; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGACGCCCTGGCAGCGCGCACTGTCATGGAACGAATGTCTGCTCGGCGCTCTGG 61
    |||||||
Db 48 CGCAGCCCTGGCAGCGCGCACTGTCATGGAACGAATTTTCTGCTCGGCGCTCTGG 107
    ||L|||||

QY 62 TGCATCCGCACTGGGTGCTGTCAGCCGCACTGTTCGAGAACTCTACACCATCGGGC 121
    ||L|||||
```

```
Db 108 TGCATCCGCACTGGGTGCTGTCAGCCGCACTGTTCGAGAACTCTACACCATCGGGC 167
QY 122 TGGGCTTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGT 170
    |||||||
Db 168 TGGGCTTNCACAGTNTTNAAGCCGACCAAGAGCCAGGAGCCAGATGGT 216

RESULT 13
AA595489/c    191 bp      mRNA      EST      18-SEP-1997
LOCUS
DEFINITION   AA595489/c    no34h04.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1102615 3',
mRNA sequence.
ACCESSION
NID          AA595489
VERSION      AA595489.1  GI:2410839
KEYWORDS
SOURCE
ORGANISM     Homo sapiens
              human.
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 191)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Apr 14, 1993 this sequence version replaced gi:692625.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.

FEATURES
    source
        1..191
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1102615"
        /clone_lib="NCI_CGAP_Pr23"
        /sex="male, pooled"
        /tissue_type="prostate tumor"
        /lab_host="SOLR (kanamycin resistant)"
        /note="Organ: prostate; Vector: Bluescript SK-; Site_1:
        EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
        Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
        GAATTCGGCAGAG 3' 3' adaptor sequence: 5'
        CTCGAGTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
BASE COUNT   43 a      28 c      57 g      63 t
ORIGIN
Query Match      13.9%; Score 161.8; DB 35; Length 191;
Best Local Similarity 95.9%; Pred. No. 9.8e-31;
Matches 163; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 998 TNCAGTCCAGCCCTCTCTCCCTCAGACCCAGCGGTCCAATGCCACTAGANTTCCCT 1057
    |||||||
Db 191 TCCAGGTACACAGCCCTCTCTCTCTCAGACCCAGCGGTCCAATGCCACTAGACTCTCCCT 132
    |||||||

QY 1058 GTACACAGTGCCTCTGTGGCANGTTGACCCAACTTACCAGTTGGTTTTCATTTT 1117
    |||||||
Db 131 GTACACAGTGCCTCTGTGGCAGTGTGACCAACCTTACCAGTTGGTTTTCATTTT 72
    |||||||

QY 1118 GTCCCTTCCCTAGATCCAGAAATAAAGTNTAAGAGAAGCGCAAAAAA 1167
    |||||||
Db 71 GTCCCTTCCCTAGATCCAGAAATAAAGTCTAAGAGAAGCGCAAAAAA 22
    |||||||
```


RESULT 14
AI385433
LOCUS
DEFINITION
AI385433 457 bp mRNA EST 27-JAN-1999
m185b05.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:473361 5' similar to gb:M22612 TRYPsinGEN I PRECURSOR
(HUMAN); gb:X04574 Mouse mRNA for preprotrypsin (MOUSE);, mRNA
sequence.
AI385433
ACCESSION
NID 94198896
VERSION AI385433.1 GI:4198896
KEYWORDS
SOURCE
EST.
house mouse.
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 457)
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
TITLE The WashU-NCI mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On May 1, 1998 this sequence version replaced gi:3105815.

Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:284105
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 putative full length read
 vector to vector length is 813
 Seq primer: -40RP from Gibco
 High quality sequence stop: 413.
 location/Qualifiers
 1..457

```

1.457
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="5 q31.3-33.2; 16"
/clone="IMAGE:473361"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGGCCGCAATTTTCTTTTCTTTTCTTTT
3']',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

| | | | | | | |
|----|--|--------|-----------------|-----------|-------------|--|
| | Query Match | 11.38; | Score 132; | DB 45; | Length 457; | |
| | Best Local Similarity | 58.5%; | Pred. No. | 3.6e-23; | | |
| | Matches 269; Conservative | 0; | Mismatches 182; | Indels 9; | Gaps 2; | |
| OY | 143 CCGACCAAGAGCCAGGAGCAGATGTGGAGGCCAGCTCTCCGTACGGCACCCAGAGT | 202 | | | | |

| | | | | |
|----|-----|---|--------|-----|
| Db | 1 | CTGTCATGAGTCTGGGGCAGCAGATGTTCCAGGGAATCAAAATCCATCCCCCACC | CGGGTT | 60 |
| QY | 203 | ACAACAGACTCTTGGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCG | | 262 |
| Db | 61 | ACTCCACCCCTGGCCACTCCCAATGACCTCATGCTCATCAAAATGAACAGAAAAATCCGTG | | 120 |
| QY | 263 | AGTCTGACACCAATCCGGAGCATCAGCAATGCTTCGAGTGGCCCTACCGGGGAACTCTT | | 322 |
| Db | 121 | ACTCTCACTCAGTGAAGCCCGTCGAATTGCTTGTGACTGTGCCACCAGGGGACTTAGT | | 180 |
| QY | 323 | GCCTCGTNTCTGGCTGGGGTCTGCTGCGCGAAGCGG-----AGAATGCCTACCGTCTGC | | 376 |
| Db | 181 | GCATGGTGTCTGGCTGGGGGACAACAGCAGCAGCCACATTACTTCCCGAAAGTCTCC | | 240 |
| QY | 377 | ACTGCGTGAACGTGTGCGTGTGCTCGAGGANGTCTGCAGTAAAGCTTATGACCCGCTGT | | 436 |
| Db | 241 | AGTGCCGTAATTAATACTGTGCTCAGTGAAGAGAGGTGTAAAACTCCTACCCAGCAGACA | | 300 |
| QY | 437 | ACCACCCAGCATGTCTGCGCCGCGGAGGGCAAGACCAGAAGACTCCTGCAACGGTG | | 496 |
| Db | 301 | TAGACAAGACCAATGTCTGCGCAGGTATG--AAGAGGGCAGGGAACTCCTGCAAGGTG | | 357 |
| QY | 497 | ACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGACAGGGCTTGTCTTTCGAAAAAG | | 556 |
| Db | 358 | ATTCCGGAGGTCCTGTGCTCTGCAATGCGCAAGTTACAGGGCTTGTCTGCGGTGAT | | 417 |
| QY | 557 | CCCCGTGTGGCCAACTTGGCGTGCACAGTGTCTACACCAA | 596 | |
| Db | 418 | TCCCGTGTGCTCAGCGGACACAGACCAGTGTCTACACCAA | 457 | |

RESULT 15
W73168
LOCUS
DEFINITION
W73168 498 bp mRNA EST 16-OCT-1996
zd55el1.r1 Soares_fetal_heart_NBHH19W Homo sapiens cDNA clone
IMAGE:344588 5' similar to SW:TR72_CANFA P06872 TRYPSINOGEN,
ANIONIC PRECURSOR ;, mRNA sequence.
W73168
ACCESSION
NID g1383322
VERSION W73168.1 GI:1383322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 498)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 5, 1995 this sequence version replaced g1:798187.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LILN; contact the
 IMAGE Consortium (info@image.liln.gov) for further information.
 Insert length: 761 Std Error: 0.00
 Seq primer: ETPprimer
 High quality sequence stop: 455.
 Location/Qualifiers
 1. .498
 /organism="Homo sapiens"
 /db_xref="GDB:1269963"
 /db_xref="taxon:9606"
 /map="4"
 /clone="IMAGE:344588"


```

/clone_lib="Soares_fetal_heart_NbHHL19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCCGACATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHHL19W."

```

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 116 a | 150 c | 128 g | 104 t |
| ORIGIN | | | | |

| | | | | |
|-----------------------|--------------|------------------|-----------------|-------------|
| Query Match | 10.18; | Score 118.2; | DB 26; | Length 498; |
| Best Local Similarity | 61.08; | Pred. No. 1e-19; | | |
| Matches 227; | Conservative | 0; | Mismatches 136; | Indels 9; |
| | | | | Gaps 2; |

[illegible]

Search completed: September 28, 1999, 11:29:46
Job time: 1867 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic : nucleic search, using sw model

Run on: September 25, 1999, 23:36:02 ; Search time 2910.9 Seconds
(without alignments)
1275.008 Million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167
Sequence: 1 GCGCAGCCCTGGCAGCGGC.....NTAAGAGAAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_cm: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vl: *
18: em_fun: *
19: em_htg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_cm: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vl: *
34: gb_htg1: *
35: gb_htg2: *
36: gb_in1: *
37: gb_in2: *
38: em_ba1: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pr4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | |
|------------|-------|-------------|--------------|--------------------|
| Result No. | Score | Query Match | Length DB ID | Description |
| 1 | 1050 | 90.0 | 1347 42 | AF113140 Homo sapi |

| | | | | | | |
|----|-------|------|--------|----|------------|---------------------|
| 2 | 552.2 | 47.3 | 5900 | 42 | AF113141 | AF113141 Homo sapi |
| 3 | 530.6 | 45.5 | 1140 | 3 | SSU076256 | U76256 Sus scrofa |
| 4 | 392 | 33.6 | 1237 | 12 | AF019979 | AF019979 Mus muscu |
| 5 | 293.6 | 25.2 | 4740 | 42 | AF135023 | AF135023 Homo sapi |
| 6 | 293.6 | 25.2 | 4385 | 42 | AF148532 | AF148532 Homo sapi |
| 7 | 202.4 | 17.3 | 35197 | 11 | AC005782 | AC005782 Homo sapi |
| 8 | 184.8 | 15.8 | 6592 | 10 | HSCTP11T5 | X90780 Homo sapien |
| 9 | 184 | 15.8 | 37785 | 9 | HUMXRC1G | L34079 Human XRC1 |
| 10 | 181.6 | 15.6 | 8174 | 5 | I34189 | I34189 Sequence 5 |
| 11 | 171.2 | 14.7 | 281 | 10 | HUMRSSA19 | K03500 Human 37 bp |
| 12 | 165.6 | 14.2 | 1665 | 9 | D87943 | D87943 Human DNA f |
| 13 | 164.6 | 14.1 | 6592 | 10 | HSCTP11T5 | X90780 Homo sapien |
| 14 | 164 | 14.1 | 205251 | 11 | AC005392 | AC005392 Homo sapi |
| 15 | 163.6 | 14.0 | 16689 | 10 | HSA011712 | AJ011712 Homo sapi |
| 16 | 160.4 | 13.7 | 42710 | 42 | AC006942 | AC006942 Homo sapi |
| 17 | 156.2 | 13.4 | 4033 | 9 | AB006136 | AB006136 Homo sapi |
| 18 | 154 | 13.2 | 16689 | 10 | HSA011712 | AJ011712 Homo sapi |
| 19 | 153 | 13.1 | 986 | 5 | A42048 | A42048 Sequence 1 |
| 20 | 153 | 13.1 | 969 | 11 | HUMSERPROT | L33404 Human strat |
| 21 | 151 | 12.9 | 35465 | 11 | AC005525 | AC005525 Homo sapi |
| 22 | 143.6 | 12.3 | 1438 | 5 | E13202 | E13202 Human gene |
| 23 | 143.6 | 12.3 | 1419 | 9 | D78203 | D78203 Homo sapien |
| 24 | 143.6 | 12.3 | 1506 | 10 | HSU62801 | U62801 Human prote |
| 25 | 143.6 | 12.3 | 1451 | 11 | AF013988 | AF013988 Homo sapi |
| 26 | 142.6 | 12.2 | 861 | 4 | AF011352 | AF011352 Petromyzo |
| 27 | 142.2 | 12.2 | 4060 | 5 | I30049 | I30049 Sequence 1 |
| 28 | 142.2 | 12.2 | 357 | 9 | HUMSATVIS | M54997 Human DNA w |
| 29 | 142.2 | 12.2 | 4067 | 10 | S51329 | S51329 AAVS1-aden |
| 30 | 141.6 | 12.1 | 37314 | 9 | HUMMDA | M63796 Human DNA f |
| 31 | 141 | 12.1 | 855 | 4 | AF011899 | AF011899 Petromyzo |
| 32 | 140.4 | 12.0 | 4776 | 11 | HSBCL3S2 | U05681 Human proto |
| 33 | 140 | 12.0 | 732 | 5 | I95869 | I95869 Sequence 2 |
| 34 | 137.8 | 11.8 | 860 | 4 | AF011898 | AF011898 Petromyzo |
| 35 | 132.6 | 11.4 | 805 | 3 | BTTRYAP | X54703 Bovine mRNA |
| 36 | 132.6 | 11.4 | 111084 | 11 | AC006486 | AC006486 Homo sapi |
| 37 | 130.8 | 11.2 | 2403 | 10 | HSNUBNDN7 | U13342 Human nucle |
| 38 | 130.2 | 11.2 | 893 | 4 | PPTRYXP | X56744 P. platessa |
| 39 | 128.2 | 11.0 | 879 | 4 | DMU58945 | U58945 Dissostichu |
| 40 | 127 | 10.9 | 978 | 4 | AF012462 | AF012462 Pleuronec |
| 41 | 126.8 | 10.9 | 1186 | 9 | AB012917 | AB012917 Homo sapi |
| 42 | 125.4 | 10.7 | 836 | 4 | AF011901 | AF011901 Petromyzo |
| 43 | 123.8 | 10.6 | 832 | 4 | AF011900 | AF011900 Petromyzo |
| 44 | 123.6 | 10.6 | 810 | 4 | SSTRYP11I | X70074 S. salar mRN |
| 45 | 123 | 10.5 | 819 | 3 | DOGTRYPA | M1589 Dog pancrea |

ALIGNMENTS

| | | | | | | |
|------------|------------|---|-----------------|------------------------------|-----|-------------|
| RESULT 1 | AF113140 | AF113140 | 1347 bp | mRNA | PRI | 25-MAR-1999 |
| LOCUS | AF113140 | Homo sapiens | serine protease | protease mRNA, complete cds. | | |
| DEFINITION | AF113140 | | | | | |
| ACCESSION | AF113140 | | | | | |
| NID | 94512029 | | | | | |
| VERSION | AF113140.1 | GI:4512029 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | | human. | | | | |
| ORGANISM | | Homo sapiens | | | | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| AUTHORS | | Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K. | | | | |
| TITLE | | Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression | | | | |
| JOURNAL | | Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999) | | | | |
| MEDLINE | | 99179024 | | | | |
| REFERENCE | | 2 (bases 1 to 1347) | | | | |
| AUTHORS | | Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K. | | | | |
| TITLE | | Direct Submission | | | | |

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA
FEATURES Location/Qualifiers
source 1.1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"

CDS
1.765
/note="androgen regulated"
/codon_start=1
/product="serine protease prostate"
/protein_id="AAD21580.1"
/db_xref="PID:g4512030"
/db_xref="GI:4512030"
/translation="MATAGNPMWMTLGLILGVAGSLVSGSQIINGEDCSPHSQPW
QAALVMENELFCGVLVHPQWVLSAHCQNSYITIGLHLEADQEPGSGMVEASLS
VRHPEYNRPLANDMLIKLDESSESPTIRISISIASQCPAGNSCLVSGWGLIANGR
MPTVLQCVNVSIVSEVCSKLYDPLVHPMFCAGGGHDQKDSNGDSGGLICNGYLO
GLVSFGKAPCGQVGPVYTNLCKTEWIEKTVQAS"
1327.1332

POLYA_site 269 a 489 c 334 g 255 t
BASE COUNT
ORIGIN

Query Match 90.0%; Score 1050; DB 42; Length 1347;
Best Local Similarity 92.4%; Pred. No. 2.4e-230;
Matches 1133; Conservative 0; Mismatches 26; Indels 67; Gaps 1;

QY 2 CGCAGCCCTGGCAGCGGCGACTGTCATGGAACGAATTGTTCTGCTCGGCGTCTGG 61
DB 122 CGCAGCCCTGGCAGCGGCGACTGTCATGGAACGAATTGTTCTGCTCGGCGTCTGG 181
QY 62 TGCATCCGAGTGGGTGCTGACGCCGACACTGTTCCAGAACTCTTACACATCGGGC 121
DB 182 TGCATCCGAGTGGGTGCTGACGCCGACACTGTTCCAGAACTCTTACACATCGGGC 241
QY 122 TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC 181
DB 242 TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC 301
QY 182 TCTCCGTACGGCACCAGAGTACAAAGACTCTGCTGCTAACGACCTCATGCTCATCA 241
DB 302 TCTCCGTACGGCACCAGAGTACAAAGACTCTGCTGCTAACGACCTCATGCTCATCA 361
QY 242 AGTGGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATGCTTCGCAGT 301
DB 362 AGTGGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATGCTTCGCAGT 421
QY 302 GCCCTACCGCGGGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
DB 422 GCCCTACCGCGGGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481
QY 362 TGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
DB 482 TGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
QY 422 TCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGG 481
DB 542 TCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGG 601
QY 482 ACTCTGCAACGGTACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 541
DB 602 ACTCTGCAACGGTACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 661
QY 542 TGTCTTTGCGAAAGCCCGGTGGCCAACTTGGCGTGCAGGTTCTTACACCACTCT 601
DB 662 TGTCTTTGCGAAAGCCCGGTGGCCAACTTGGCGTGCAGGTTCTTACACCACTCT 721
QY 602 GCAAAITCACTGAGTGATAGAGAAAACCGTCCAGNCCAGTTAATCTTGGGGAGCTGG 661
DB 722 GCAAAITCACTGAGTGATAGAGAAAACCGTCCAGGCGCACTTAATCTTGGGGAGCTGG 781

QY 662 CCCATGAATGACCCCAATATACATCTGCGGAANGAATTTCAGGAATATCTGTCCAG 721
DB 782 CCCATGAATGACCCCAATATACATCTGCGGAANGAATTTCAGGAATATCTGTCCAG 841
QY 722 CCCCTCTCTCCAGAGCCAGAGTCCAGGCCCCCAGCCCTCTCTCTCAACCAAGGG 781
DB 842 CCCCTCTCTCCAGAGCCAGAGTCCAGGCCCCCAGCCCTCTCTCTCAACCAAGGG 901
QY 782 TACAGATCC----- 791
DB 902 TACAGATCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCCCAGCCCTCTCTC 961
QY 791 -----CCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCCCAGCC 834
DB 962 CCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCCCAGCC 1021
QY 835 CCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCC 894
DB 1022 CCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCC 1081
QY 895 CCCAGCCCTCTCTCTCTCAGACCCAGAGGTTGAGGCCCCCAACCCCTCTCTCTCAGAGT 954
DB 1082 CCCAGCCCTCTCTCTCTCAGACCCAGAGGTTGAGGCCCCCAACCCCTCTCTCTCAGAGT 1141
QY 955 CAGAGGTTCAAGCCCAACCCCTCTCTCTCTCAGACCCAGAGGTTGAGGCCCCCT 1014
DB 1142 CAGAGGTTCAAGCCCAACCCCTCTCTCTCTCAGACCCAGAGGTTGAGGCCCCCT 1201
QY 1015 CCTCTCTCAGACCCAGAGGTTCAATGACCTAGANTTCCCTGTACACAGTGGCCCTT 1074
DB 1202 CCTCTCTCAGACCCAGAGGTTCAATGACCTAGANTTCCCTGTACACAGTGGCCCTT 1261
QY 1075 GTGCGANGTGACCAACCTTACAGTGTGTTTTCATTTTGTCCCTTCCCTAGAT 1134
DB 1262 GTGCGANGTGACCAACCTTACAGTGTGTTTTCATTTTGTCCCTTCCCTAGAT 1321
QY 1135 CCAGAAATAAGTNTAAGAGAAGCGC 1160
DB 1322 CCAGAAATAAGTNTAAGAGAAGCGC 1347

RESULT 2

AF113141 5900 bp DNA PRI 25-MAR-1999
LOCUS Homo sapiens serine protease prostate gene, complete cds.
DEFINITION AF113141
ACCESSION AF113141
NID 94512031
VERSION AF113141.1 GI:4512031
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 5900)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.
TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 5900)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA

FEATURES
source 1.5900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"

| | | |
|---------------------------|--|---|
| mRNA | | join(<1411. .1471,2735. .2897,3319. .3569,3653. .3789, 5062. .5817) |
| exon | | /product-"serine protease prostate" 1411. .1471 /number=1 join(1411. .1471,2735. .2897,3319. .3569,3653. .3789, 5062. .5214) /note="androgen regulated" |
| CDS | | /codon_start=1 /product-"serine protease prostate" /protein_id="AAD21581.1" /db_xref="PID:g4512032" /db_xref="GI:4512032" /translation="MATAGNPMGWFLGLILGVAGSLVSGSCQIINGEDCSPHSQPW OALVMENEFCSGVLVHPQWVLSAHCFOFNSYTIIGLHSLLEADQEPGSOMEASLS VRHPEYNRPPLANDMLIKDESVESEDTRISIASQCPTAGNSCLVSGWGLLANGR MPTVLOCVNVSIVSEVSESKLYDPLVHPSMFCAGGGHDQKDCSNGDSGPIICNGYLQ GLVSFGKAPCGOVGPVGYTTLCKFTWEIEKTVOAS" |
| exon | | 2735. .2897 /number=2 3319. .3569 |
| exon | | /number=3 3653. .3789 |
| exon | | /number=4 5062. .5817 |
| BASE COUNT | 1417 a 1643 c 1569 g 1271 t | |
| ORIGIN | | |
| Query Match | 47.3%; Score 552.2; DB 42; Length 5900; | |
| Best Local Similarity | 87.8%; Pred. No. 1e-116; | |
| Matches 634; Conservative | 0; Mismatches 21; Indels 67; Gaps 1; | |
| QY 491 | ACGGTGACTCTGGGGGGCCCCCTGATCTGCACACGGGTACTTCGAGGGCCCTGTGCTTTTCG | 550 |
| Db 5060 | AGGGTGACTCTGGGGGGCCCCCTGATCTGCACACGGGTACTTCGAGGGCCCTGTGCTTTTCG | 5119 |
| QY 551 | GAAAAGCCCCGTGTGGGCCAACTGGCGGTGCCAGGTGTCTACACCACCTCTGCAAAATTCA | 610 |
| Db 5120 | GAAAAGCCCCGTGTGGGCCAACTGGCGGTGCCAGGTGTCTACACCACCTCTGCAAAATTCA | 5179 |
| QY 611 | CTGAGTGATAGAAAAACCGTCCAGNCAGTTAACTCTGGGCACTGGGAACCATGAAA | 670 |
| Db 5180 | CTGAGTGATAGAAAAACCGTCCAGNCAGTTAACTCTGGGCACTGGGAACCATGAAA | 5239 |
| QY 671 | TTCACCCCCCAATACATCCTTCGCGGAANGAATTACAGGAATATCTGTCCACAGCCCCTCTC | 730 |
| Db 5240 | TTCACCCCCCAATACATCCTTCGCGGAANGAATTACAGGAATATCTGTCCACAGCCCCTCTC | 5299 |
| QY 731 | CCTCAGGCCCAGGAGTCCAGGCCCCCAGGCCCTCTCTCCCTCAAACAAGGGGTACAGATCC | 790 |
| Db 5300 | CCTCAGGCCCAGGAGTCCAGGCCCCCAGGCCCTCTCTCCCTCAAACAAGGGGTACAGATCC | 5359 |
| QY 791 | ----- | 791 |
| Db 5360 | CCAGCCCCCTCCTCCTCAGAACCCAGGAGTCCAGACCCCCCAGCCCCCTCCTCCTCAGACC | 5419 |
| QY 791 | -----CCAGCCCCCTCCTCCTCCTCAGAACCCAGGAGTCCAGACCCCCCAGCCCCCTCCTCN | 843 |
| Db 5420 | CAGGAGTCCAGGCCCTCCTCCTCAGAACCCAGGAGTCCAGACCCCCCAGCCCCCTCCTCC | 5479 |
| QY 844 | TCAGACCCAGGAGTCCAGGCCCTCCTCCTCNTCAGACGCAGGAGTCCAGACCCCCCAGCCCN | 903 |
| Db 5480 | TCAGACCCAGGAGTCCAGGCCCTCCTCCTCNTCAGACGCAGGAGTCCAGACCCCCCAGCCCN | 5539 |
| QY 904 | TCNTCCGTGAGACCCAGGGGTGACGGCCCCCAACCCCTCNTCNTCAGAGTCCAGAGGTCC | 963 |
| Db 5540 | TCCTCCCTCAGACCCAGGGGTGACGGCCCCCAACCCCTCNTCNTCAGAGTCCAGAGGTCC | 5599 |
| QY 964 | AAGCCCCCAACCCCTCGTTCCTCCCAAGACCCAGAGGTTCAGAGTCCAGCCCTCCTCCTCA | 1023 |
| Db 5600 | AGCCCCCAACCCCTCGTTCCTCCCAAGACCCAGAGGTTCAGAGTCCAGCCCTCCTCCTCA | 5659 |

```

QY 1024 GACCCAGCGGTCACATGCCACCTAGANTNTCCCTGTACACAGTGCCCCCTGTGGCANGT 1083
      |||||||
Db 5660 GACCCAGCGGTCACATGCCACCTAGACTCTCCCTGTACACAGTGCCCCCTGTGGCAGGT 5719
      |||||||

QY 1084 TGACCCACACCTTACAGTGGTTTTTCATTTTTCCTTCCCTTAGATCCAGAAATA 1143
      |||||||
Db 5720 TGACCCACACCTTACAGTGGTTTTTCATTTTTCCTTCCCTTAGATCCAGAAATA 5779
      |||||||

QY 1144 AA 1145
      ||
Db 5780 AA 5781

RESULT 3
SSU76256 LOCUS SSU76256 1140 bp mRNA MAM 03-JAN-1998
DEFINITION Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,
complete cds.
ACCESSION U76256
NID 92737920
VERSION U76256.1 GI:2737920
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Cetartiodactyla; Suidae; Sus.
TITLE 1 (bases 1 to 1140)
Simmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,
Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
Purification, Characterization and Cloning of Enamel Matrix Serine
Proteinase 1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1140)
AUTHORS Simmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,
Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA
FEATURES
source location/Qualifiers
1. 1140
/organism="Sus scrofa"
/db_xref="taxon:9823"
69. 140
/feature="the signal peptide is cleaved after Ala24 and the
preprotein is secreted into the developing enamel matrix"
69. 833
/feature="EMSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAB94638.1"
/db_xref="PID:92737921"
/db_xref="GI:2737921"
/translation="MTAAGNPWGWFGLLGLTGSIAFINGGHIINGEDCNPHSQPW
QAALFLEDDFFCGGLVHPQWVLSAHCQNSYTIIGLHNLPEQEPGSMELASLS
IQHPENPSMANDIMLIKLESVSLDTRNISVVSQCPTPGDSCLVSGWRLASGR
LPQVLOCVNISVASEVCKARPYVHPSPMFCAGGQDKSCHSGSGPLICNGSLQ
GLVSGQAQCGQPNVPGVYTNLCKFTDWIQTITQAS"
159. 830
/feature="the preprotein is cleaved following His30
activating the serine proteinase; the active protein has
an apparent molecular weight of 34 kDa and a derived
molecular weight of 24 kDa"
/product="enamel matrix serine proteinase 1"
/order(177. 179,549. 551)
/feature="disulfide bond"
/order(234. 236,282. 284)
/feature="disulfide bond"
join(279. 281,414. 416,687. 689)
/feature="encodes catalytic triad"
405. 407
/feature="encodes potential glycosylation location; yields a

```


misc_feature blank cycle during protein sequencing"
order(408.410,789.791)
/note="disulfide bond"
misc_feature order(510.512,705.707)
/note="disulfide bond"
misc_feature order(675.677,750.752)
/note="disulfide bond"
BASE COUNT 245 a 366 c 295 g 234 t
ORIGIN

Query Match 45.5%; Score 530.6; DB 3; Length 1140;
Best Local Similarity 81.3%; Pred. No. 9.7e-112;
Matches 625; Conservative 0; Mismatches 143; Indels 1; Gaps 1;

OY 2 CGCAGCCCTGGCAGCGGCACATGGAAGAATGTTCTGCTGGGGCTCCTGG 61
DB 190 CGCAGCCCTGGCAGCGGCACATGTTCTGGAAGACGACTTTTCTGCGGGGGCTCTGG 249
OY 62 TGCATCCGACAGTGGGTGCTGTACCCGACACTGTTCCAGAAGTCTTACACCATCGGGC 121
DB 250 TGCATCCGACATGGGTGCTGTACCCGACACTGTTCCAAAATTCCTACACCATCGGGC 309
OY 122 TGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGTTGAGGCGCAGCC 181
DB 310 TGGGCTGCACATCTTGAGCGCGAACAAGACCCGGCAGCCAGATGATGAGGCGCAGCC 369
OY 182 TCTCCGTACGGCACCAGAGTACAAAGACTCTTGTCTGCTAAGCAGCCTCATGCTCATCA 241
DB 370 TCTCCATTCAGCACCAGAAATCAACAGCCGCTCTATGSCCAACGACCTCATGCTCATCA 429
OY 242 AGTTGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATGCTTCGAGT 301
DB 430 AGTTGAAGAATCGGTGTGCTGTGACACCGTCCGGAACATCAGCGTCTCTCCAGT 489
OY 302 GCCCTACCGCGGGGGAACCTCTGCTCTGNTCTGGCTGGGCTCTGCTGGCGAAGCGCAGAA 361
DB 490 GCCCGACCCCTGGGATCTTGCTGCTGTTCCGGCTGGGCTGGGCTGGCGAAGCGCAGAA 549
OY 362 TGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
DB 550 TGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609
OY 422 TCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGCGAAGCAGAAAGG 481
DB 610 GCTACGGCCCGTGTACACCCAGCATGTTCTGCGCTGGCGGAGGCGAAGCAGAAAGG 669
OY 482 ACTGCTGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
DB 670 ACTGCTGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
OY 542 TGTCTTTCGAAAAAGCCCGTGTGCGCAACTTGGCGTGGCAGGTGTCTACACCAACTCT 601
DB 730 TGTCTTTCGAAAAAGCCCGTGTGCGCAACTTGGCGTGGCAGGTGTCTACACCAACTCT 789
OY 602 GCAAAATTCAGTGTGATAGAAAAACCGTCCAGNCCAGTTAATCTGGGGAGTGGGAA 661
DB 790 GCAAAATTCAGTGTGATAGAAAAACCGTCCAGNCCAGTTAATCTGGGGAGTGGGAA 849
OY 662 CCCATGAATTCAGTGTGATAGAAAAACCGTCCAGNCCAGTTAATCTGGGGAGTGGGAA 721
DB 850 GCAATGAATTCAGTGTGATAGAAAAACCGTCCAGNCCAGTTAATCTGGGGAGTGGGAA 909
OY 722 -CCCTCTCTCTCAGGCGCCAGAGTCCAGCCCGCCAGCCCTCTCTCC 769
DB 910 TGCTCTCTCTCTCAGGCGCCAGAGTCCAGCCCGCCAGCCCTCTCTCC 958

RESULT 4
AF019979 1237 bp mRNA ROD 01-JAN-1999
LOCUS AF019979 1237 bp mRNA ROD 01-JAN-1999
DEFINITION Mus musculus enamel matrix serine proteinase 1 precursor, mRNA,
complete cds.

ACCESSION AF019979
NID 94090846
VERSION AF019979.1 GI:4090846
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1237)
AUTHORS Simmer,J.
TITLE Enamel Matrix Serine Proteinase 1 (EMSP1)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1237)
AUTHORS Simmer,J.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA

FEATURES
source location/Qualifiers
1.1237
/organism="Mus musculus"
/strain="Swiss Webster"
/db_xref="taxon:10090"
/cell_type="ameloblast"
52.126
52.819
sig_peptide
CDS
/function="degradation of the enamel matrix during enamel
maturation"
/note="EMSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAC98894.1"
/db_xref="PID:94090847"
/db_xref="GI:4090847"

misc_feature
127.144
/note="encodes propeptide"
mat_peptide
145.816
/product="matrix serine proteinase 1"
BASE COUNT 289 a 386 c 286 g 276 t
ORIGIN

Query Match 33.6%; Score 392; DB 12; Length 1237;
Best Local Similarity 65.8%; Pred. No. 4.3e-80;
Matches 628; Conservative 0; Mismatches 311; Indels 16; Gaps 4;

OY 2 CGCAGCCCTGGCAGCGGCACATGTCATGGAAGAATGTTCTGCTGGGGCTCCTGG 61
DB 176 CGCAGCCCTGGCAGCGGCACATGTTCTCAGAGACGGTTTCTGCTGGGAGTCTTGG 235
OY 62 TGCATCCGACAGTGGGTGCTGTACAGCCGACACTGTTCCAGAACTCTTACACCATCGGGC 121
DB 236 TGCATCCGACAGTGGGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 295
OY 122 TGGGCTGCACAGTCTTGAGGCGCGACCAAGAGCCAGGAGCCAGATGTTGAGGCGCAGCC 181
DB 296 TGGGCTGCATTAACCTGAAGGCTCCCAAGAGCCTGGCAGCCGAGTGTAGAGGCCACC 355
OY 182 TCTCCGTACGGCACCAGAGTACAAAGACTCTGCTGCTAAGACCTCATGCTCATCA 241
DB 356 TCTCCATCCAGCACCACCACTCAAGATCCTTCTTTCGAACGATCATGCTCATCA 415
OY 242 AGTTGACGAATCCGTGTCCAGTGTGACACCAATCCGAGCATCAGCATGCTTCGAGT 301
DB 416 AACTGAACGAGTCAGTGATAGAGTCTAACACATTCAGAGACATCCCTGTGGCTAACCAAT 475
OY 302 GCCCTACCGCGGGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361

Db 476 GCCCCACTCTGGAGATACCTGCCTAGTCTGTGTTGGGTCACCTAAGAATGGAAAC 535
QY 362 TGCCCTACCGTGTGCACCTGCGTGAACGTGCGTGTGTCTGAGGANGTCTGACGTAAAC 421
Db 536 TGCCAGCCTCTCTGCAGTGTGTGAATCTCTAGTGGCGTCAGAGAGACTTGCCGGCTGC 595
QY 422 TCTATGACCCGCTGTACACCCAGCAGTCTTGCGCCGGCGGAGGACAGACAGAAAG 481
Db 596 TGTATGACCCCTGTATACCACTCAGTATGTCTGCGCTGAGAGGAGACAGACCAAAAGC 655
QY 482 ACTCCTGCAACGCTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGACAGGGCTTG 541
Db 656 ACTCCTGCAATGGGGACTCTGGGGGTCCCATAGTGTGCAACAGATCCCTACAAGGGCTCG 715
QY 542 TGTCTTGGGAAAAGCCCGTGTGGCCAATTGGCGTGCAGGTGTCTACACCAACCTCT 601
Db 716 TGTCTATGGACAAGAAAGTGTGGCGACCTGGCATACCAAGTGTCTACACCAATCTCT 775
QY 602 GCAAAATTCAGTGTGATAGAGAAAACCGTCCAGNCCAGTTAACTCTGGGACTGGGA 661
Db 776 GCAAGTTCACTAAGTGTGATCCAGACCAATCATTCAGACAACCTAGCTATG-----A 826
QY 662 CCCATGAATTCAGCCCAATATACATCCTGCGGAANGAATTCAGGAATATCTGTCCAG 721
Db 827 CTGACTACAGGCCCCACATGAGCCCTGTGAGGGAACCTGAGACTCAGGATTTGCTCTAG 886
QY 722 CCCCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCCAGCCCTCTCTCTCAAAACCAAGGG 781
Db 887 -CCCTTCTCCTCAGATCTACAAGGTGAGGCCCTAGCCTCTCTCTCAGACTCACAAGTTC 945
QY 782 TACAGATCCCGAGCCCTCTCTCTCAGACCCAGAGTCCAGAC-CCCCAGCCCTCTNT 840
Db 946 CAGA-----CTCCAATCTCTCTCTCAGACCCACCAAGTCCAGACTCCAATTTCTCTCTC 1000
QY 841 CCNTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACGACGAGTCCAGACCCCCAGC 900
Db 1001 AGATCCAAGATTTCAGACCAATATACCTCTTTCAGATTCAGGGATCCATGCCACAACCT 1060
QY 901 CCNTCCTCCGTGAGACCCAGGGTGCAGGCCCCCAACCCCTCTCTCTCAGAGTTC 955
Db 1061 CCTCCCTGGAACACAGAGTGAACCCACCTGCATTTCTCTGTGTGACAGTGTTC 1115

RESULT 5

AF135023 4740 bp DNA PRI 20-APR-1999
LOCUS AF135023 Homo sapiens kallikrein-like protein 1 KLK-L1 gene, partial cds.
DEFINITION AF135023
ACCESSION AF135023
NID 94589272
VERSION AF135023.1 GI:4589272
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 4740)
AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.
TITLE Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4

JOURNAL Blochem. Biophys. Res. Commun. (1999) In press

REFERENCE 2 (bases 1 to 4740)

AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.

TITLE Direct Submission

JOURNAL Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount

Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,

Canada

FEATURES location/Qualifiers

SOURCE 1. 4740

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13.3-q13.4"

join(<2263. .2425,2847. .3097,3181. .3317,4588. .>4740)

/product="kallikrein-like protein 1 KLK-L1"
join(<2263. .2425,2847. .3097,3181. .3317,4588. .4740)
/codon_start=3

/product="kallikrein-like protein 1 KLK-L1"

/protein_id="AAD26424.1"

/db_xref="GI:4589273"

/db_xref="GI:4589273"

/translation="SLVSGSCQIINGEDCSPHSQPMQALVMENELFCSGVLVHPQVLSAHCFO NSTYITGLHLSLEADQEPGDSOMVEASLSVRHPEYNRPPLANDMLIKLDESSES D TIRSIISISQCP T AGNSCLVSGWGLANGRMPTVLCQVNSVSEVCSKLYDPLIHP SMFCAGGGQDKDSCN GDSGFP LICNGYLOGLVSRKAPCGQVGPVYITNLCKFT EMI EKTVOAS"

BASE COUNT 1106 a 1275 c 1308 g 1051 t

ORIGIN

Query Match 25.2%; Score 293.6; DB 42; Length 4740;
Best Local Similarity 81.4%; Pred. No. 1.2e-57;
Matches 389; Conservative 0; Mismatches 6; Indels 83; Gaps 1;

QY 99 CCAGAACTCCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGGCCAGG 158

Db 2841 CCACAGCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGGCCAGG 2900

QY 159 GAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCACAGAGTACACACAGACTTGTCT 218

Db 2901 GAGCCAGATGCTGAGGCCAGCCTCTCCGTACGGCACCACAGAGTACACACAGACTTGTCT 2960

QY 219 CGCTAACGACCTCATGCTCATCAAGTTGAGCGAATCCGTGTCCAGTCTGACACCATCCG 278

Db 2961 CGCTAACGACCTCATGCTCATCAAGTTGAGCGAATCCGTGTCCAGTCTGACACCATCCG 3020

QY 279 GAGCATCAGCATTTGCTTGGCAGTCCCTACCGCGGGGAACCTTTGCTCTGTTCTGGCTG 338

Db 3021 GAGCATCAGCATTTGCTTGGCAGTCCCTACCGCGGGGAACCTTTGCTCTGTTCTGGCTG 3080

QY 339 GGGTCTGTGGCGAAC----- 355

Db 3081 GGGTCTGTGGCGAACGGGTGAGCTCACGGGTGTGTCTGCCCTCTTCAAGAGAGTCTC 3140

QY 355 -----GGCAGAAATGCCCTACCGTGTG 375

Db 3141 TGCCAGATGCGGGGGCTGACCCAGAGCTCTGCGTCCAGGACAGAAATGCCCTACCGTGTG 3200

QY 376 CACTGCGTGAACGTGTGCGTGTCTGAGGANGTCTGACGTAGACTCTATGACCCGCTG 435

Db 3201 CAGTGCCTGAACGTGTGCGTGTCTGAGGAGGTCTGCAAGTCTATGACCCGCTG 3260

QY 436 TACCACCCAGCATGTTCTGCGCGCGGCGAGGGCAAGACCAAGAGACTCTCTGCAAG 493

Db 3261 TACCACCCAGCATGTTCTGCGCGCGGCGAGGGCAAGACCAAGAGACTCTCTGCAAG 3318

RESULT 6

AF148532 4385 bp DNA PRI 09-JUN-1999
LOCUS AF148532 Homo sapiens kallikrein 4 (KLK4) gene, complete cds.
DEFINITION AF148532
ACCESSION AF148532
NID 95020095
VERSION AF148532.1 GI:5020095
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 4385)
AUTHORS Stephenson,S.A., Verity,K., Ashworth,L. and Clements,J.A.
TITLE Localization of a new prostate specific antigen-related serine

protease gene, KLK4, is evidence for an expanded human kallikrein

(KLK) gene family cluster on chromosome 19q13.3-13.4

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4385)

AUTHORS Stephenson,S.A., Ashworth,L. and Clements,J.A.

TITLE Direct Submission
JOURNAL Submitted (03-MAY-1999) School of Life Science, Queensland
University of Technology, George Street, Brisbane, Queensland 4001,
Australia
FEATURES
source
1.4385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.3-q13.4"
/join("<551..611,1875..2037,2459..2709,2793..2929,
4200..>4352)
/gene="KLK4"
/product="Kallikrein 4"
<551..>4352
/gene="KLK4"
/join(551..611,1875..2037,2459..2709,2793..2929,4200..4352)
/gene="KLK4"
/function="serine protease"
/codon_start=1
/product="kallikrein 4"
/protein_id="AAD38019.1"
/db_xref="PID:95020096"
/db_xref="GI:5020096"
/translation="MATAGNPMGFWLGLYLGVAGSLVSGSCSQIINGEDCSPHSQPW
QALVWENELFCSGVLVHPQWYSAHCFQNSYTIIGLHSLADQEPGSGQWVEASLS
VRHPEYNRPPLANDMLIKDEYSESDTIRISISIASQCPTAGNSCLVSGWELLANGR
MPTVLQCVNVSVSEVCSKLDPLHPSMFCAGGGDQDKDSCNGDSGGPLICNGYLO
GLVSFKAPCGGVGPVYTNLCKFTWIEKTVQAS"
BASE COUNT 1013 a 1184 c 1192 g 996 t
ORIGIN
Query Match 25.2%; Score 293.6; DB 42; Length 4385;
Best Local Similarity 81.4%; Pred. No. 1.2e-57;
Matches 389; Conservative 0; Mismatches 6; Indels 83; Gaps 1;
QY 99 CCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGG 158
Db 2453 CCACAGCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGG 2512
QY 159 GAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACTTGTGCT 218
Db 2513 GAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACTTGTGCT 2572
QY 219 CGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCG 278
Db 2573 CGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCG 2632
QY 279 GAGCATCAGCATTTGCTTCGACAGTGCCTACCGGGGGAACCTTGCCTGCTNTCTGCTG 338
Db 2633 GAGCATCAGCATTTGCTTCGACAGTGCCTACCGGGGGAACCTTGCCTGCTNTCTGCTG 2692
QY 339 GGGTCTGCTGCGAAC----- 355
Db 2693 GGGTCTGCTGCGAACGGGTGAGCTCAGCGGTGTGTCTGCCCTCTCAAGGAGGTCTCTC 2752
QY 355 -----GGCAGAATGCCCTAACCGTGCTG 375
Db 2753 TGCCCAAGTCGGGGGGCTGACCCAGAGCTTCGCTCCACGAGAAATGCTTACCGTGCTG 2812
QY 376 CACTGCGTGAACGTGTGGTGCTGTCTGAGGANGTCTGCAGTAAGCTTATGACCCGCTG 435
Db 2813 CAGTGCGTGAACGTGTGGTGCTGTCTGAGGAGGTCTGAGTAAGCTTATGACCCGCTG 2872
QY 436 TACCACCCAGCATGTCTGCGCGGCGAGGCGAGACCAAGAGACTCTGCAACG 493
Db 2873 TACCACCCAGCATGTCTGCGCGGCGGAGGCGAGACCAAGAGACTCTGCAACG 2930
RESULT 7
AC005782
LOCUS - AC005782 35197 bp DNA PRI 05-OCT-1998

DEFINITION Homo sapiens chromosome 19, cosmid R31855, complete sequence.
ACCESSION AC005782
NID 93702296
VERSION AC005782.1 GI:3702296
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 35197)
Lamerdin,J.E., MCCready,P.M., Skowronski,E., Adamson,A.W.,
Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Ganes,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Altix,C., Andreise,T., Frankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of an ~700 kb region in 19q13.4 between D19S268
and SYT3
TITLE Unpublished.
JOURNAL 2 (bases 1 to 35197)
REFERENCE Lamerdin,J.E.
AUTHORS Direct Submission
TITLE Submitted (05-OCT-1998) Joint Genome Institute, Lawrence Livermore
JOURNAL National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
FEATURES
source
1.35197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R31855"
/chromosome="19"
/map="19q13.4 between D19S268 and SYT3"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19- specific cosmid
library"
/note="Cosmid library LL19NC03 was constructed at LLNL
from flow-sorted chromosomes from human-hamster hybrid
5HL2-B which carries chromosome 19 as its only human
chromosome."
148..255
/note="BLASTX similarity to Q08509 (311..346); match:
0.52, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir||S39983
eps8 protein - mouse >gi|309217 (L21671) Eps8 [Mus]"
155..243
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 98.000"
251..279
/rpt_family="GC_rich"
338..430
/note="BLASTX similarity to Q08509 (343..373); match:
0.51, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir||S39983
eps8 protein - mouse >gi|309217 (L21671) Eps8 [Mus]"
338..412
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
517..660
/note="BLASTX similarity to Q08509 (368..415); match:
0.52, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir||S39983
eps8 protein - mouse >gi|309217 (L21671) Eps8 [Mus]"
517..665
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000--DDS
similarity to overlapping ESTs:-(611..665) R34158
yh83h09.s1 Homo sapiens cdna clone 136385 3', similar to
SP:S39983 S39983 EPS8 PROTEIN; (1..55); 100%
identity.--(621..665) R80263 y196h06.s1 Homo sapiens cdna
clone 147131 3'; (1..45); 100% identity."
complement(1145..1258)
/rpt_family="FLAM_C"

misc_feature 1441..1582
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 75.000--DDS
similarity to overlapping ESTs:-R34158 yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (56..194); 94% identity.--R80263 y196h06.s1 Homo sapiens cDNA clone 147131 3'; (46..186); 94% identity.--H13945 yj08d11.s1 Homo sapiens cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (21..161); 94% identity."
complement(1621..1652)
/rpt_family="GC_rich"
1679..1707
/note="DDS similarity to overlapping ESTs:-R34158 yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (195..223); 97% identity.--R80263 y196h06.s1 Homo sapiens cDNA clone 147131 3'; (187..215); 97% identity.--H13945 yj08d11.s1 Homo sapiens cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (162..190); 93% identity."
complement(1880..1975)
/rpt_family="MSR1"
3160..3240
/rpt_family="MIR"
complement(3241..3527)
/rpt_family="AluSc"
3607..3702
/rpt_family="MIR"
3904..4036
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 99.000--DDS
similarity to overlapping ESTs:-R3904..3988) R80263 y196h06.s1 Homo sapiens cDNA clone 147131 3'; (216..300); 96% identity.--(3904..4036) H13945 yj08d11.s1 Homo sapiens cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (191..323); 95% identity."
3989..4036
/note="DDS similarity to R34158 yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (224..271); 100% identity."
4124..4255
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 90.000--DDS
similarity to :-(4124..4215) R34158 yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (272..364); 94% identity.--(4124..4136) H13945 yj08d11.s1 Homo sapiens cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN; (324..336); 100% identity."
4455..4581
/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 95.000"
4777..4951
/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 93.000"
complement(4817..4878)
/rpt_family="GC_rich"
5366..5498
/note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: excellent, score: 93.000--DDS
similarity to:-(5367..5498) H13944 yj08d11.r1 Homo sapiens cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8 PROTEIN; (368..336); 93% identity."
5591..5677
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 89.000--DDS
similarity to:-(5591..5825) H13944 yj08d11.r1 Homo sapiens cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8 PROTEIN; (235..1); 99% identity."
complement(5619..5988)
/note="BLASTN similarity to A1126400 (11..380); match: 0.99, score: 5.0e-144; database searched: month.na; qc55f11.xl Soares_placenta_8to9weeks_2NbHP8to9W Homo

misc_feature sapiens cDNA clone IMAGE:1713549 3', mRNA sequence [Homo sapiens]
complement(5691..5961)
/note="DDS similarity to R34157 yh83h09.r1 Homo sapiens cDNA clone 136385 5'. Score: 513 Identity: 268/270 (99%)."
5852..5983
/note="BLASTN similarity to D25793 (1..132); match: 0.98, score: 2.8e-45; database searched: est; Human colon 3'directed MboI cDNA, HUMGS04161, clone cm1029."
complement(6867..6897)
/rpt_family="GC_rich"
7429..8057
/rpt_family="LINE2"
complement(8297..8602)
/rpt_family="AluSx"
8979..9294
/note="DDS similarity to T23821 seq1081 Homo sapiens cDNA clone b4HB3MA-Cot51.5-HAP-Ft-23 3'; Score: 597 Identity: 311/319 (97%).--Additional EST matches: A1125316, T33476 and many others"
complement(9335..9621)
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 92.000"
complement(9677..9763)
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 100.000"
complement(10285..10389)
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 100.000"
complement(10580..10651)
/rpt_family="GC_rich"
complement(10685..10725)
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 85.000"
complement(10810..10931)
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 79.000--(10928..10906) DDS similarity to R35625 yg66g06.r1 Homo sapiens cDNA clone 38310 5'; (319..341); 78% identity."
complement(11079..11139)
/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 83.000--DDS
similarity to R35625 yg66g06.r1 Homo sapiens cDNA clone 38310 5'; (254..318); 91% identity."
complement(11216..11296)
/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 100.000--DDS
similarity to overlapping ESTs:-R20513 y946h05.r1 Homo sapiens cDNA clone 35856 5'; (193..271); 99% identity.--R35625 yg66g06.r1 Homo sapiens cDNA clone 38310 5'; (173..253); 98% identity."
misc_feature
Query Match 17.3%; Score 202.4; DB 11; Length 35197;
Best Local Similarity 82.2%; Pred. No. 6.8e-37;
Matches 278; Conservative 0; Mismatches 49; Indels 11; Gaps 4;
OY 712 CTGTTCCAGCCCCCTCTCCTCAGGCCAGGAGTCAGGCCCCCAGCCCTCTCCCTC 771
DB 12059 CAGGCCCGAGCCCTCTCCTCAGACCAGGAGTCAGACCTCAGCCCTCTCCCTC 12118
OY 772 AAACCAAGGCTACAGATCCCGACCCCTCTCCTCAGACCAGGAGTCAGACCCCA 831
DB 12119 AGACCCAGAGTCCAGGCCCGACCCCTCTCCTCAGACCAGGAGCCCA 12176
OY 832 GCCCTCTCCTCAGACCCAGAGT-----CCAGCCCTCTCCTCAGACGACGAGA 884
DB 12177 GCCCTCTCCTCAGACCCAGAGTCCAGGCCCGACCCCTCTCCTCAGACGAGA 12236
OY 885 GTCCAGACCCCGAGCCCTCCTCCTCAGACCCAGGAGTGCAGGCCCCCAACCCCTC 944
DB 12237 GTCCAGA-CCTCCAGCCCTCTCTCCTCAGACCCAGGAGTCCAGG-CGCCAGCCCTCT 12294

QY 1041 CCACCTAGANTTCCT 1057
Db 3304 CCCAGCCCTCTCCT 3320

RESULT 9

LOCUS HUMXRC1G/c 37785 bp DNA PRI 30-JAN-1995
DEFINITION Human XRC1 DNA repair gene, genomic.
ACCESSION L34079
NID 9642116
VERSION L34079.1 GI:642116
KEYWORDS Alu repeat; DNA repair protein; tandem satellite array.
SOURCE Homo sapiens (tissue library: LL19NC02-F2) adult blood DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37785)
AUTHORS Lamerdin,J.E., Carrano,A.V., Thompson,L.H., Montgomery,M.A.,
Stilwagen,S.A., Scheidecker,L. and Tebbs,R.S.
TITLE Genomic sequence comparison of the human and mouse XRC1 DNA repair
gene regions
JOURNAL Genomics (1995) In press
FEATURES
source location/Qualifiers
1..37785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="UV5HL9-5B"
/cell_type="lymphocyte"
/dev_stage="adult"
/tissue_type="blood"
/tissue_lib="LL19NC02-F2"
/map="19q13.2"
4268..36195
/gene="XRC1"
4268..4318
/partial
/gene="XRC1"
/note="exon 1; G00-120-737"
4724..4816
/gene="XRC1"
/note="exon 2; G00-120-737"
18706..18816
/partial
/gene="XRC1"
/note="exon 3; G00-120-737"
24922..25080
/partial
/gene="XRC1"
/note="exon 4; G00-120-737"
26043..26117
/partial
/gene="XRC1"
/note="exon 5; G00-120-737"
26214..26325
/partial
/gene="XRC1"
/note="exon 6; G00-120-737"
complement(26342..26484)
/partial
/gene="XRC1"
/note="human chromosome 19-specific tandem repeat, pE670"
26635..26744
/partial
/gene="XRC1"
/note="exon 7; G00-120-737"
26818..26929
/partial
/gene="XRC1"
/note="exon 8; G00-120-737"
complement(27008..27410)
/partial
/gene="XRC1"
satellite

exon /note="human chromosome 19-specific tandem repeat, pE670"
27451..27709
/partial
/gene="XRC1"
/note="exon 9; G00-120-737"
28039..28155
/partial
/gene="XRC1"
/note="exon 10; G00-120-737"
32609..32702
/partial
/gene="XRC1"
/note="exon 11; G00-120-737"
32859..32991
/partial
/gene="XRC1"
/note="exon 12; G00-120-737"
complement(33001..33360)
/partial
/gene="XRC1"
/note="human chromosome 19-specific tandem repeat, pE670"
33480..33534
/partial
/gene="XRC1"
/note="exon 13; G00-120-737"
33629..33768
/partial
/gene="XRC1"
/note="exon 14; G00-120-737"
35353..35443
/partial
/gene="XRC1"
/note="exon 15; G00-120-737"
35899..35975
/partial
/gene="XRC1"
/note="exon 16; G00-120-737"
36082..36195
/partial
/gene="XRC1"
/note="exon 17; G00-120-737"
BASE COUNT 8717 a 9329 c 9806 g 9929 t 4 others
ORIGIN

Query Match 15.8%; Score 184; DB 9; Length 37785;
Best local Similarity 73.3%; Pred. No. 1.1e-32;
Matches 297; Conservative 0; Mismatches 96; Indels 12; Gaps 5;

QY 712 CTGTTCCAGCCCTCTCCTCAGGCCAGGAGTCAGGCCCCCAGCCCTCTCCTC 771
Db 27335 CAGGCCAGCCCTCTCCTCAGGCCAGGAGTCAGGCCCCCAGCCCTCTCCTC 27276

QY 772 AAACCAAG-GTACAGATCCAGCCCTCTCCTCAGACCAGAGTCAGACCCCC 830
Db 27275 AGACCCAGAGTCAGGCCCCCAGCCCTCTCCTCAGGCCAGAGTCAG-GCCCC 27217

QY 831 AGCCCTCTCCTCAGACCAGAGTCAG-----CCCTCTCTCCTCAGAGCAG 882
Db 27216 AGCCCTCTCCTCAGACCAGGAGTCAGGCCCAACCCCTCTCTCAGACCAG 27157

QY 883 GAGTCAGACCCCCAGCCCTCTCCTCAGACCAGGGGTGAGGCCCAACCCCTC 942
Db 27156 GGGTCAG-GCTCCAGCCCTCTCTCTCAGACCAGGGGTTAGGCTCAGCCCTC 27098

QY 943 NTCCNTCAGAGTCAGAGGTCAGACCCCAACCCCTCTCTCCCAAGACCAGAGTNCAG 1002
Db 27097 CTCCCTCGGATTCAGGAATCCAGGCCCCAGCCCTCTCTCTCCTCAGACCAGGGGTCAG 27038

QY 1003 G-TCCAGACCCCTCTCTCTCAGACCAGCGGTCATGCCACCTAGANTTCCCTGTAC 1061
Db 27037 GCCCCAGCCCTCTCTCTCAGATTCAGAGCAGAGGCTCAGCCCGCTTCCCTAGG 26978

QY 1062 ACAGTCCCCCTGTGGCANGTTGACCCCAACCTTACCACTTGTT 1106
Db 26977 ACACAGAGACACAGGCCCTGTGTGAACAAGGATCTAGTAGCT 26933

RESULT 10

LOCUS 134189 8174 bp DNA PAT 27-JAN-1997
DEFINITION Sequence 5 from patent US 5595900.
ACCESSION 134189
NID 91824980
VERSION 134189.1 GI:1824980
KEYWORDS
SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 8174)
Lowe,J.B.
TITLE

Methods and products for the synthesis of oligosaccharide
structures on glycoproteins, glycolipids, or as free molecules, and
for the isolation of cloned genetic sequences that determine these
structures
Patent: US 5595900-A 5 21-JAN-1997;
Location/Qualifiers
1. 8174

JOURNAL
FEATURES

source
BASE COUNT 1628 a 2229 c 2322 g 1995 t
ORIGIN

Query Match 15.6%; Score 181.6; DB 5; Length 8174;
Best Local Similarity 77.3%; Pred. No. 4.2e-32;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

QY 700 ATTCAGGAATATCTGTTCCAGCCCCCTCCTCCCTCAGGCCCAAGAGTCCAGGCCCCAGC 759
Db 1025 ACTTAGGAGTCCAGGCTCCGCGCCCTCCTCCTCAGACCAGAGTCCAGGCCCTGC 966

QY 760 CCTCTCCTCCTCAAAACCAAGG-GTACAGATCCCAAGCCCCCTCCTCCTCAGACCAGAG 818
Db 965 CCTCTCCTCCTCAGACCAGAGTCCAGACCAGGCCCTCCTCCTCAGACCAGAG 906

QY 819 TCCAGACCCCCAGCCCTCCTCCTCCTCAGACCAGAGTCCAG-----CCCTCCTC 870
Db 905 TCCAGG-TCCCTAGCCCTACTCCTCCTCAGACCAGAGTCCAGACCAAGCTCCTCCTC 847

QY 871 CMTACAGCAGGAGTCCAGACCCCCAGCCCTCCTCCTCAGACCAGAGGAGTCCAGAGC 930
Db 846 CCTCAGACCCAGGAGGCCCAAGTTCCTCCTCCTCAGATCCAGAGTACAGG- 787

QY 931 CCCCAACCCCTCCTCCTCCTCAGAGTCCAGAGTCCAGGCCCAACCCCTCCTCCTCAGAG 990
Db 787 -CCAGACCTCCTCCTCCTCAGACCCAGAGTCCAGGCCCAACCCCTCCTCCTCAGAG 729

QY 991 CCAGAGTNCAGTCCAGCCCTCCTCCTCCTCAGACCAGCGGTCCAA 1038
Db 728 CCAGAGTCCAGAGCCCTCCTCCTCCTCCTCAGACAGAGGCTA 681

RESULT 11

LOCUS HUMRSSA19 281 bp DNA PRI 09-MAY-1996
DEFINITION Human 37 bp minisatellite repeats, specific to chromosome 19.
ACCESSION K03500
NID 9337672
VERSION K03500.1 GI:337672
KEYWORDS repeat region.
SOURCE Homo sapiens (clone library: COS 4C) DNA.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 281)
AUTHORS... Das,H.K., Jackson,C.L., Miller,D.A., Leff,T. and Breslow,J.L.

TITLE The human apolipoprotein C-II gene sequence contains a novel
JOURNAL J. Biol. Chem. 262 (10), 4787-4793 (1987)
MEDLINE 87165892
FEATURES
source Location/Qualifiers
1. 281
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="COS 4C"
/cell_line="JY"
1. 37
/note="sat rpt copy A"
38. 75
/note="sat rpt copy B"
76. 113
/note="sat rpt copy C"
114. 151
/note="sat rpt copy D"
152. 188
/note="sat rpt copy E"
189. 225
/note="sat rpt copy F"
226. 262
/note="sat rpt copy G"
263. 281
/note="sat rpt partial copy H"

BASE COUNT 58 a 147 c 44 g 32 t
ORIGIN Chromosome 19, long arm.

Query Match 14.7%; Score 171.2; DB 10; Length 281;
Best Local Similarity 81.2%; Pred. No. 1.3e-29;
Matches 221; Conservative 0; Mismatches 41; Indels 10; Gaps 2;

QY 781 GTACAGATCCCAAGCCCTCCTCCTCAGACCAGAGTCCAGACCCAGCCCTCCT 840
Db 2 GTCAAGACCCCAAGCCCTCCTCCTCCTCAGACCAGAGTCAAGAACCCCAAGCCCTCCT 61

QY 841 CMTACAGACCCAGAGT-----CCAGCCCTCCTCCTCCTCAGACGACGAGTCCAGA 891
Db 62 CCTCAGACCCAGAGTCAAGAACCCCAAGCCCTCCTCCTCAGACCAGAGTCAAGA 121

QY 892 CCCCCAGCCCTCCTCCTCAGACCAGAGGAGTCCAGGCCCAACCCCTCCTCCTCAG 951
Db 122 CCCCCAGCCCTCCTCCTCCTCAGACTCATGATCCAGACCCCAAGCCCTCCTCCTCAG 181

QY 952 AGTCAGAGTCCAAAGCCCAACCCCTCCTCCTCAGACCAGAGTNCAG-GTCCAGAG 1010
Db 182 ACCCAGAGTCCAGACCCCAAGCCCTCCTCCTCAGACCAGAGTCCAGACCCCAAG 241

QY 1011 CCTCTCCTCCTCAGACCAGCGGTCCAATGCC 1042
Db 242 CCTCTCCTCCTCAGACCAGAGTCCAGGCC 273

RESULT 12

LOCUS D87943 1665 bp DNA PRI 29-JUL-1997
DEFINITION Human DNA for alpha(1,2)fucosyltransferase, intron 1.
ACCESSION D87943
NID 91842173
VERSION D87943.1 GI:1842173
KEYWORDS FUT1; alpha (1, 2) fucosyltransferase.
SOURCE Homo sapiens peripheral leukocytes DNA.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1665)
AUTHORS Kimura,H.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hiroshi Kimura, Kurume University, School of Medicine, Department of Legal Medicine; 67 Asahimachi, Kurume, Fukuoka 830, Japan

(E-mail:hkimura@kurume.ktarn.or.jp, Tel:0942-31-7554,
Fax:0942-31-7700)

REFERENCE
AUTHORS
TITLE

Koda,Y., Soejima,M. and Kimura,H.
Structure and expression of H-type GDP-L-fucose:beta-D-galactoside
2-alpha-L-fucosyltransferase gene (FUT1). Two transcription start
sites and alternative splicing generate several forms of FUT1 mRNA
J. Biol. Chem. 272 (11), 7501-7505 (1997)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE

Larsen,R.D., Ernst,L.K., Nair,R.P. and Lowe,J.B.
Molecular cloning, sequence, and expression of a human
GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase cDNA
that can form the H blood group antigen
Proc. Natl. Acad. Sci. U.S.A. 87 (17), 6674-6678 (1990)

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Kelly,R.J., Ernst,L.K., Larsen,R.D., Bryant,J.G., Robinson,J.S. and
Lowe,J.B.
Molecular basis for H blood group deficiency in Bombay (Oh) and
para-Bombay individuals
Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5843-5847 (1994)

COMMENT
FEATURES

Sequence updated (10-Feb-1997) by:Hiroshi Kimura.
Location/Qualifiers
1..1665

gene
1..1665
/gene="FUT1"
Intron
1..1665
/gene="FUT1"
/number=1

BASE COUNT 349 a 382 c 513 g 421 t
ORIGIN

Query Match 14.2%; Score 165.6; DB 9; Length 1665;
Best Local Similarity 75.8%; Pred. No. 2.1e-28;
Matches 279; Conservative 0; Mismatches 74; Indels 15; Gaps 6;

OY 700 ATTCAGGATATCTGTTCCAGCCCTCTCTCCCTCAGGCCAGGAGTCCAGCCCCAGC 759

DB 381 ACTTAGAGTCCAGGCTCCCGCCCTCTCTCCCTCAGACCAGAGTCCAGCCCCCTGC 322

OY 760 CCCTCTCTCCCTCAACAAGG-GTACAGATCCCAAGCCCTCTCTCTCAGACCAGAG 818

DB 321 CCTCTCTCTCAGACCCAGAGAGTCCAGAGCCCAAGCCCTCTCTCTCAGACCAGAG 262

OY 819 TCCAGACCCCCAGCCCTCTCTCTCAGACCAGAGTCCAG-----CCCTCTCTC 870

DB 261 TCCAG-GTCCCTAGCCCTACTCTCTCAGACCCAGAGTCCAGACCAAGCTCTCTCTC 203

OY 871 CNTCAGACGACGAGTCCAGACCCCCAGCCCTCTCTCTCAGAGCCAGGGGTGACAGC 930

DB 202 CCTCAGACCCAGAGGCCCAAGTCCCAAGCCCTCTCTCTCTCAGATCCAGAGTACAGG- 143

OY 931 CCCCACCCCTCTCTCTCAGAGTCCAGAGTCCAGACCCCAAGCCCTCTCTCTCTCTCTC 990

DB 143 -CCAGCCCTCTCTCTCTCAGATCCAGAGTACAGG--CCAGACCTCTCTCTCTCTCTC 87

OY 991 CCAGAGTNCAGG-TCCAGCCCTCTCTCTCTCAGACCCAGCGGTCCATGCACTAGA 1049

DB 86 CCAGAGTCCAGGCCCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCAAGCCC 27

OY 1050 NTNTCCCT 1057

DB 26 TCCTCCCT 19

RESULT 13

HSCPTIT5/c
LOCUS

DEFINITION
ACCESSION
NID

HSCPTIT5 6592 bp DNA PRI 27-APR-1999
Homo sapiens TNNT3 gene.
X90780 X90781 X90782
94071059
X90780.1 GI:4071059
cardiac troponin I; TNNT3 gene.
human.

ORGANISM

REFERENCE

1 (bases 1 to 6592)
Bhavsar,P.R., Brand,N.J., Yacoub,M.H. and Barton,P.J.R.
Isolation and characterization of the human cardiac troponin I gene
(TNNT3)
Genomics 35 (1), 11-23 (1996)

JOURNAL
MEDLINE
REFERENCE

2 (bases 1 to 6592)
Barton,P.J.R., Cullen,M.E., Townsend,P.J., Brand,N.J., Mullen,A.J.,
Norman,D.A.M., Bhavsar,P.R. and Yacoub,M.H.
Close physical linkage of human troponin genes: organization,
sequence, and expression of the locus encoding cardiac troponin I
and slow skeletal troponin T
Unpublished
3 (bases 1 to 6592)
Bhavsar,P.R.
Direct Submission
Submitted (15-AUG-1995) P.K. Bhavsar, Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science,Tech.&
Med., Dovehouse Street, London SW3 6LY, UK
Revised by [3]
4 (bases 1 to 6592)
Barton,P.J.R.
Direct Submission
Submitted (22-DEC-1998) Barton P.J.R., Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science,Tech.&
Med., Dovehouse Street, London SW3 6LY, UK

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

COMMENT

This sequence is directly followed by the slow skeletal troponin T
gene (AC AJ011712, AJ011713)
Related sequence M64247.
Location/Qualifiers
1..6592

FEATURES

source

gene

exon

mRNA

TATA_signal

prim_transcript

5'UTR

CDS

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone_id="P1"
/map="p13.3-q13.3"
/note="putative"
join(1107..1260,1532..1544,1706..1789,2195..2236,
2507..2638,4009..4098,4633..4809,6442..6525)
/gene="TNNT3"
1107..6592
/gene="TNNT3"
1107..1249
/gene="TNNT3"
/number=1
/gene="TNNT3"
join(1250..1260,1532..1544,1706..1789,2195..2236,
2507..2638,4009..4098,4633..4809,6442..6525)
/gene="TNNT3"
/product="cardiac troponin I"
/protein_id="CAA62301.1"
/db_xref="PID:e196058"
/db_xref="PID:g1524066"
/db_xref="GI:1524066"
/db_xref="SWISS-PROT:P19429"


```
/translation="MADSSDAAREPRAPAPIRRRSSNYRAYATEPHAKKSKISAS
RKLQKTLTLLQIAKQELEREAEERGERGALSTRQPLELAGLGFALQDLRQJHA
RVDKVEERYDIEAKVTKNITEIADLTQKIFDLRGKFKRPTLRVRISADAMQALLG
ARAKESLDLRHLKQVKKEDTEKENREVGWDRKNIDALSGMEGRKKKFFES"
exon      1532. 1544
           /gene="TNNI3"
           /number=2
exon      1706. 1789
           /gene="TNNI3"
           /number=3
exon      2195. 2236
           /gene="TNNI3"
           /number=4
exon      2507. 2638
           /gene="TNNI3"
           /number=5
exon      4009. 4098
           /gene="TNNI3"
           /number=6
exon      4633. 4809
           /gene="TNNI3"
           /number=7
exon      6442. 6592
           /gene="TNNI3"
           /number=8
3'UTR      6526. 6592
           /gene="TNNI3"
           /gene="TNNI3"
polyA_signal 6568. 6573
           /gene="TNNI3"
           /note="putative"

BASE COUNT 1424 a 1836 c 2048 g 1284 t
ORIGIN
```

Query Match 14.1%; Score 164.6; DB 10; Length 6592;
Best Local Similarity 70.1%; Pred. No. 3.3e-28;
Matches 260; Conservative 0; Mismatches 99; Indels 12; Gaps 3;

```
OY 703 CAGGAATATCTGTCCAGAGCCCTCCTCCCTCAGGCGCCAGAGTCCAGGCCCGAGCCCC 762
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 765 CAAGAGTCCAGGCCCCCAGGCCCTCCTCCTCAGACCCAGAGTCCAGGCCCGAGCTCC 706

OY 763 TCCTCCCTCAACCAAGGTTACAGATCCCGAGCCCTCCTCCTCAGACCCAGAGTCCA 822
    || ||||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 705 TCTTCCCTCAGACCCAGAGTCCAGGCCCCCAGGCCCTCCTCCTCAGACCCAGAGTCCA 646

OY 823 GACCCCCAGCCCTCCTCCTCAGACCCAGAGT-----CCAGCCCTCCTCCTCCTC 874
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 645 G--GCCCCAGCCCTCCTCCTCAGACCCAGAGTCCAGGCCCCCAGGCCCTCCTCCTC 588

OY 875 AGACGCGAGGAGTCCAGACCCCGAGCCCTCCTCCTCAGACCCAGGCGGTGAGGCCCCC 934
    ||||| ||||| ||||| ||||| || | | | | | | | | | | | | | | | | | | |
DB 587 AGACCCAGGGGTTCCAGGCCCCCAGGCCCTCCTCCTCAGACCAAGGAGTCCAGGCGGCC 528

OY 935 A--ACCCCTCCTCCTCCTCAGAGTCCAGAGTCCAGGCCCGAGCCCTCCTCCTCAGACCC 992
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 527 AGCAGACTTCTCTCTCAGACTCAGAGTCTGCACCTCAGCTCCTCCTCCTCAGACCA 468

OY 993 AGAGGTNCAGGTCCAGGCCCTCCTCCTCAGACCCAGGCGGTCCATGACACCTAGANTN 1052
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 467 GGAGTCCAGGAGCCCAATCCCTCTCCCTCAGACCCGAGAGTCCAGGACCCCAATCCCTC 408

OY 1053 TCCCTGTACAC 1063
    ||||| | | |
DB 407 TCCCTCAGAC 397
```

RESULT 14
AC005392/c AC005392 205251 bp DNA PRI 04-DEC-1998
DEFINITION Homo sapiens chromosome 19, CIT-HSP BAC 490g23 (BC338531), complete
sequence.
AC005392 AC005392

```
NID 93399669
VERSION AC005392.1 GI:3399669
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 205251)
AUTHORS Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Krommiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
clustered CEA/PSG gene family
TITLE Unpublished
JOURNAL 2 (bases 1 to 205251)
REFERENCE Lamerdin,J.E.
AUTHORS Direct Submission
TITLE Submitted (06-AUG-1998) Joint Genome Institute, Lawrence Livermore
JOURNAL National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from q centromere to telomere. BAC 338531
(CIT-HSP 490g23) is adjacent to cosmid F24498 to the left and
overlaps BAC 52850 (CIT-HSP 6117) to the right by approximately 35
kb. Accession does not represent the sequence of the entire BAC
clone insert at this time. Additional chr 19 map and sequence
information are available at:
http://www-bio.lnl.gov/genome/genome.html.
FEATURES
source
Location/Qualifiers
1..205251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="490g23"
/chromosome="19"
/map="19q13.2 between BCKDHA and D19S217"
/clone_1lb="CIT-HSP"
/tissue_type="sperm"
/note="LNL clone name: BC338531"
complement(566..1150)
/rpt_family="MER67C"
1154..1359
/rpt_family="AluJo"
1360..1398
/rpt_family="AluJo"
complement(1399..1483)
/rpt_family="MER67C"
2634..2937
/rpt_family="Alusx"
3399..3495
/rpt_family="MLT1A1"
complement(3674..3716)
/rpt_family="MIR"
3807..4109
/rpt_family="Alusq"
4129..4417
/rpt_family="Alusq"
4498..4578
/rpt_family="Alusq"
/rpt_family="AluJ/FRAM"
4581..4774
/rpt_family="L1MC/D"
complement(5020..5087)
/rpt_family="Alu"
complement(5095..5281)
/rpt_family="L1"
complement(6416..6557)
/rpt_family="L1M4"
complement(6558..6860)
/rpt_family="AluY"
complement(7329..7457)
/rpt_family="MER4_internal"
```



```

repeat_region 7478. .7839 /rpt_family="MER50"
repeat_region complement(7869. .7920)
repeat_region /rpt_family="(TA)n"
repeat_region 7937. .8213
repeat_region /rpt_family="MER4_internal"
repeat_region complement(8259. .9153)
repeat_region /rpt_family="LIPA2"
repeat_region 9684. .9975
repeat_region /rpt_family="AluJo"
repeat_region 11171. .11311
repeat_region /rpt_family="MIR"
repeat_region complement(12081. .12261)
repeat_region /rpt_family="MIR"
repeat_region 12362. .12397
repeat_region /rpt_family="MLTID"
repeat_region 12399. .12691
repeat_region /rpt_family="AluSg"
repeat_region 12692. .12869
repeat_region /rpt_family="MLTID"
repeat_region complement(13008. .13038)
repeat_region /rpt_family="AluSg"
repeat_region complement(13074. .13218)
repeat_region /rpt_family="MER11A"
repeat_region complement(13256. .13442)
repeat_region /rpt_family="MER11A"
repeat_region complement(13622. .13993)
repeat_region /rpt_family="HERVK"
repeat_region complement(14138. .14301)
repeat_region /rpt_family="AluSg"
repeat_region complement(14308. .14692)
repeat_region /rpt_family="MER57_internal"
repeat_region 14855. .15152
repeat_region /rpt_family="AluSp"
repeat_region complement(15179. .16657)
repeat_region /rpt_family="MER65_internal"
repeat_region complement(16026. .16426)
repeat_region /note="DDS similarity to overlapping ESTs:-(16426. .16026)
repeat_region W43007 zc24g09.sl Soares senescent fibroblasts NbHSF Homo
repeat_region sapiens cDNA clone 323296 3', similar to gb:M20882
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN E PRECURSOR
repeat_region (HUMAN); (401. .10); 99% identity.--(16494. .16238) W49630
repeat_region zc42h12.r1 Soares senescent fibroblasts NbHSF Homo sapiens
repeat_region cDNA clone 325031 5' similar to gb:M23575
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN C PRECURSOR
repeat_region (HUMAN); (268. .524); 99% identity."
repeat_region complement(16667. .16961)
repeat_region /rpt_family="AluY"
repeat_region complement(16963. .17056)
repeat_region /rpt_family="MER65_internal"
repeat_region 17252. .17402
repeat_region /rpt_family="MER5B"
repeat_region complement(18346. .18385)
repeat_region /note="DDS similarity to W43007 zc24g09.sl Soares
repeat_region senescent fibroblasts NbHSF Homo sapiens cDNA clone 323296
repeat_region 3', similar to gb:M20882 PREGNANCY-SPECIFIC BETA-1
repeat_region GLYCOPROTEIN E PRECURSOR (HUMAN); (441. .402); 100%
repeat_region identity."
repeat_region complement(18346. .18429)
repeat_region /note="DDS similarity to W49630 zc42h12.r1 Soares
repeat_region senescent fibroblasts NbHSF Homo sapiens cDNA clone 325031
repeat_region 5', similar to gb:M23575 PREGNANCY-SPECIFIC BETA-1
repeat_region GLYCOPROTEIN C PRECURSOR (HUMAN); (184. .267); 100%
repeat_region identity."
repeat_region 18772. .19186
repeat_region /rpt_family="LTR16B"
repeat_region complement(19312. .19498)
repeat_region /rpt_family="MIR"
repeat_region complement(19577. .19678)
repeat_region /rpt_family="MIR"
repeat_region 19679. .19978
repeat_region /rpt_family="L1"
repeat_region 19981. .20294

```

[illegible]

Db 201461 TCCGACCCAGAGTCCAGACCCCGAGCCCTCCTCCTCAGACCCAGAGATCCAGGCCCC 201402

RESULT 15
LOCUS HSA011712/c 16689 bp DNA PRI 27-APR-1999
DEFINITION Homo sapiens TNNT1 gene, exons 1-11 (and joined CDS).
ACCESSION AJ011712
NID g4056561
VERSION AJ011712.1 GI:4056561
KEYWORDS slow skeletal; TNNT1 gene; troponin T.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 16689)
AUTHORS Barton,P.J., Cullen,M.E., Townsend,P.J., Brand,N.J., Mullen,A.J.,
Norman,D.A., Bhavsar,P.K. and Yacoub,M.H.
TITLE Close physical linkage of human troponin genes: organization,
sequence, and expression of the locus encoding cardiac troponin I
and slow skeletal troponin T
JOURNAL Genomics 57 (1), 102-109 (1999)
MEDLINE 99208666
REFERENCE 2 (bases 1 to 16689)
AUTHORS Barton,P.J.R.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1998) Barton P.J.R., Cardiothoracic Surgery,
National Heart & Lung Institute, Dovehouse Street, London SW3 6JB,
United Kingdom
REMARK Revised on 20-DEC-1998
COMMENT Related sequences : X90780 (cardiac troponin I gene, directly
upstream)
AJ011713 (TNNT1 gene, exons 12-14).
FEATURES
source location/Qualifiers
1..16689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="DMPC-HFF#1-1495-E10"
/map="q13.4"
1..2580
/gene="TNNT1"
/note="promoter region"
1..16689
/gene="TNNT1"
2552..2557
/gene="TNNT1"
/note="putative TATA box"
prim_transcript join(2581..16689,AJ011713.1:1..1846)
exon 2581..2626
/gene="TNNT1"
/number=1
join(2581..2626,4587..4629,4734..4747,5048..5074,
6189..6210,9829..9892,10447..10563,10789..10866,
13679..13792,14577..14686,AJ011713.1:519..609,
AJ011713.1:748..788,AJ011713.1:1715..1846)
/note="alternatively spliced mRNA"
/product="troponin T, slow skeletal muscle"
join(2581..2626,4587..4629,4734..4747,5048..5074,
5289..5321,6189..6210,9829..9892,10447..10563,
10789..10866,13679..13792,14577..14686,
AJ011713.1:471..609,AJ011713.1:748..788,
AJ011713.1:1715..1846)
/note="alternatively spliced mRNA"
mRNA join(2581..2626,4587..4629,4734..4747,5048..5074,
5289..5321,6189..6210,9829..9892,10447..10563,
10789..10866,13679..13792,14577..14686,
AJ011713.1:519..609,AJ011713.1:748..788,
AJ011713.1:1715..1846)
/note="alternatively spliced mRNA"
/product="troponin T, skeletal and cardiac muscle"
join(2581..2626,4587..4597)
5'UTR ..

exon
/gene="TNNT1"
4587..4629
/gene="TNNT1"
/number=2
join(4598..4629,4734..4747,5048..5074,5289..5321,
6189..6210,9829..9892,10447..10563,10789..10866,
13679..13792,14577..14686,AJ011713.1:471..609,
AJ011713.1:748..788,AJ011713.1:1715..1760)
/note="alternatively spliced isoform"
/codon_start=1
/product="slow skeletal muscle troponin T"
/protein_id="CAA09752.1"
/db_xref="PID:e1361228"
/db_xref="PID:g4056564"
/db_xref="GI:4056564"
/translation="MSDTEEQYEYEEQPEEEAAEEEEAPEPEPVAPEPEERPKPSR
PVPPLIPPKIPEGERVDFDDIHRKRMKDLLEQLTLDVHFEQRKKEEELVALKER
IERRRSERAEQORFTEKERERQAKLAEEKMRKEEEAKKRAEDDAKKKYLNMGAH
FGGYLYKAEQKRGKRGKRGKRVRLSERKKPLDIDYMGEEQLRKAQELSDWIHOLE
AREKAQELSDWIHOLESEKFDLMAKLRKQKYEINVLNRIASHAQFRKAGKGRVGR
WK"
CDS join(4598..4629,4734..4747,5048..5074,5289..5321,
6189..6210,9829..9892,10447..10563,10789..10866,
13679..13792,14577..14686,AJ011713.1:519..609,
AJ011713.1:748..788,AJ011713.1:1715..1760)
/note="alternatively spliced isoform"
/codon_start=1
/product="slow skeletal muscle troponin T"
/protein_id="CAA09751.1"
/db_xref="PID:e1361227"
/db_xref="PID:g4056563"
/db_xref="GI:4056563"
/translation="MSDTEEQYEYEEQPEEEAAEEEEAPEPEPVAPEPEERPKPSR
PVPPLIPPKIPEGERVDFDDIHRKRMKDLLEQLTLDVHFEQRKKEEELVALKER
IERRRSERAEQORFTEKERERQAKLAEEKMRKEEEAKKRAEDDAKKKYLNMGAH
FGGYLYKAEQKRGKRGKRGKRVRLSERKKPLDIDYMGEEQLRKAQELSDWIHOLE
SEKFDLMAKLRKQKYEINVLNRIASHAQFRKAGKGRVGRWK"
CDS join(4598..4629,4734..4747,5048..5074,6189..6210,
9829..9892,10447..10563,10789..10866,13679..13792,
14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,
AJ011713.1:1715..1760)
/note="alternatively spliced isoform"
/codon_start=1
/product="slow skeletal muscle troponin T"
/protein_id="CAA09750.1"
/db_xref="PID:e1361226"
/db_xref="PID:g4056562"
/db_xref="GI:4056562"
/translation="MSDTEEQYEYEEQPEEEAAEEEEERPKPSRPVVPPLIPPKI
PEGERVDFDDIHRKRMKDLLEQLTLDVHFEQRKKEEELVALKERIERRRSERAEQ
ORFTEKERERQAKLAEEKMRKEEEAKKRAEDDAKKKYLNMGAHFGGYLYKAEQK
RGKRGKRGKRVRLSERKKPLDIDYMGEEQLRKAQELSDWIHOLESEKFDLMAKLR
KQKYEINVLNRIASHAQFRKAGKGRVGRWK"
4734..4747
/gene="TNNT1"
/number=3
5048..5074
/gene="TNNT1"
/number=4
5289..5321
/gene="TNNT1"
/note="alternative exon"
/number=5
6189..6210
/gene="TNNT1"
/number=6
9829..9892
/gene="TNNT1"
/number=7
10447..10563
/gene="TNNT1"
/number=8
10789..10866
exon
exon
exon
exon
exon
exon
exon


```

exon      /gene="TNNT1"
           /number=9
           13679..13792
exon      /gene="TNNT1"
           /number=10
           14577..14686
exon      /gene="TNNT1"
           /number=11
           14687..>16689
Intron

```

```

/gene="TNNT1"
/note="approx. 200 bp gap with AJ011713"
/number=11

```

[illegible]

| | | | | |
|---------------------------|--------|--------------------|-----------|---------------|
| Query Match | 14.0%; | Score 163.6; | DB 10; | Length 16689; |
| Best Local Similarity | 72.3%; | Pred. No. 5.2e-28; | | |
| Matches 258; Conservative | 0; | Mismatches 95; | Indels 4; | Gaps 4; |

| | | | |
|----|-------|--|-------|
| Qy | 703 | CAGGAATATCTGTTCCAGCCCCCTCTCTCCCTCAGGCCAGGATTCAGGCCCCAGCCCC | 762 |
| | | | |
| Db | 14257 | CAGGATCCAGACCCCGAGCCCCCTCTCTCTCAGACCCAGGATTCAGGCCCCGAGCCCC | 14198 |

Qy 763 TCCTCCCTCAACCAAGG-GTACAGATCCCCAGCCCCCTCCTCCCTCAGACCCAGGAGTCC 821
|||||
Db 14197 TCCTCCCTCAGACCCAGGAGTGCAGGCCCCCAGCCCCCTCCTCCCTCAGACCCAGGAGTCC 14138
|||||

Oy 822 AGACCCCCCAGCCC-CTCNTCCNTCAGACCAGGAGTCCAGGCCCTCTCCNTCAGACGC 880
||| ||| | ||| ||| ||| |
Db 14137 AGGCCCCCAGCCCTCTCTCAGACTGATTCACAAGCCCCAGACCCCTCTCGATTAGAAC 14078

QY 881 AGAGTCCAGACCCCCCAGCCCTCCTCGTCAGACCAGGGGTGCAGGCCCAACCCC 940
| | | | | | | | | | | | | | | | | | | | | |
Db 14077 CAGAAGTCG-GCCCCAGCCCTCTTTTCAGACCCAGGGGTTCAGGCCCTCAGCCCC 14019

Qy 941 TCNTCCNTCAGAGTCAGAGGTTCCAAGCCCCCAACCCCTCGTGTCCCAGAGACCAGAGGTC 1000
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14018 TCCTCCCTCAG -CCAGGAGTCCAGGACCCCAAGCCCTCCTCCNTCAGAACCCAGGCCTCC 13960

| | | | |
|----|-------|---|-------|
| Qy | 1001 | AGGTCCACGCCCTCTCCCTCAGACCCAGCGGTCCATGCGCACCTAGANTNTCCCT | 1057 |
| | | | |
| Db | 13959 | AGACCCACGCTCCTCTCCCTCAGACCCAGGAGTTACAGGGCCACGCCCTCTCCCT | 13903 |

Search completed: September 25, 1999, 23:36:44
Job time: 7347 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:57:00 ; Search time 273 Seconds
(without alignments)
1069.502 Million cell updates/sec

Title: US-09-030-606-175
Perfect score: 1167
Sequence: 1 GCGCAGCCCTGGCAGCGCGC.....NTAGAGAACGCCAATAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|--------|-----------------------|
| 1 | 1151 | 98.6 | 1167 | 1 | V58647 | Prostate tumour sp |
| 2 | 1151 | 98.6 | 1167 | 1 | V61252 | CDNA sequence of p |
| 3 | 1111.2 | 95.2 | 1248 | 1 | V58644 | Prostate tumour sp |
| 4 | 1111.2 | 95.2 | 1248 | 1 | V61249 | CDNA sequence of p |
| 5 | 1036.2 | 88.8 | 1265 | 1 | V58645 | Prostate tumour sp |
| 6 | 1036.2 | 88.8 | 1265 | 1 | V61250 | CDNA sequence of p |
| 7 | 950.2 | 81.4 | 1386 | 1 | V11855 | CDNA sequence of p |
| 8 | 721.4 | 61.8 | 871 | 1 | V37495 | Homo sapiens Tub I |
| 9 | 352.2 | 30.2 | 1119 | 1 | V58648 | Prostate prostate-spe |
| 10 | 352.2 | 30.2 | 1119 | 1 | V61253 | CDNA sequence of p |
| 11 | 331.4 | 28.4 | 402 | 1 | X41114 | Human secreted pro |
| 12 | 292.8 | 25.1 | 1459 | 1 | V58646 | Prostate tumour sp |
| 13 | 292.8 | 25.1 | 1459 | 1 | V61251 | CDNA sequence of p |
| 14 | 228.8 | 19.6 | 234 | 1 | V58522 | Prostate tumour sp |
| 15 | 228.8 | 19.6 | 234 | 1 | V61168 | CDNA sequence of p |
| 16 | 188.4 | 16.1 | 1476 | 1 | X16295 | Human kallikrein e |
| 17 | 181.6 | 15.6 | 8174 | 1 | Q13332 | GDP-Fuc:beta-D-gal |
| 18 | 181.6 | 15.6 | 8174 | 1 | Q56908 | DNA encoding a gly |
| 19 | 181.6 | 15.6 | 8174 | 1 | T61677 | Human alpha(1,2)-f |
| 20 | 154.6 | 13.2 | 1089 | 1 | T39783 | Human amyloid prec |
| 21 | 153 | 13.1 | 986 | 1 | O81203 | Human stratum corn |
| 22 | 143.6 | 12.3 | 1438 | 1 | T79126 | Human serine prote |
| 23 | 143.6 | 12.3 | 1526 | 1 | V07152 | Protease M, a nove |
| 24 | 140 | 12.0 | 732 | 1 | O53487 | DNA encoding Zyme |
| 25 | 137.4 | 11.8 | 4067 | 1 | O63192 | Human adeno-associ |
| 26 | 126 | 10.8 | 1146 | 1 | V84589 | Human secreted pro |
| 27 | 125.4 | 10.7 | 833 | 1 | V42925 | DNA encoding a hum |
| 28 | 124.6 | 10.7 | 693 | 1 | V61861 | A. contortrix prot |
| 29 | 111.8 | 9.6 | 741 | 1 | N81633 | Human spleen tryps |
| 30 | 111.8 | 9.6 | 744 | 1 | T03999 | Human pancreatic t |
| 31 | 111.8 | 9.6 | 744 | 1 | T04000 | Human pancreatic t |
| 32 | 111.8 | 9.6 | 744 | 1 | T04001 | Human pancreatic t |
| 33 | 110.6 | 9.5 | 1333 | 1 | T48519 | Human neurotrophin |
| 34 | 110.6 | 9.5 | 1333 | 1 | T63251 | Mouse neurotrophin |
| 35 | 110.2 | 9.4 | 790 | 1 | V24548 | Trypsinogen-like p |
| 36 | 107.8 | 9.2 | 683 | 1 | O63794 | Trypsinogen |
| 37 | 106.2 | 9.1 | 701 | 1 | O63795 | Bovine trypsin gen |
| 38 | 100.8 | 8.6 | 897 | 1 | T49878 | Porcine trypsinoge |
| 39 | 99.4 | 8.5 | 701 | 1 | X23298 | Human TRVI trypsin |
| 40 | 98 | 8.4 | 944 | 1 | V84052 | Nucleic acid encod |
| 41 | 95.8 | 8.2 | 957 | 1 | O20501 | Encodes fibrinogen |
| 42 | 90.6 | 7.8 | 1003 | 1 | T13316 | Korean Viper Salmo |
| 43 | 88.2 | 7.6 | 925 | 1 | N93196 | Monkey recombinant |

C 44 87.4 7.5 90 1 V58593 Prostate tumour sp
C 45 87.4 7.5 90 1 V61208 CDNA sequence of p

ALIGNMENTS

RESULT 1
ID V58647 standard; cDNA: 1167 BP.
AC V58647;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
FH Key
FT CDS Location/Qualifiers
FT 28..645
FT /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
DR P-PSDB; W69388.
PT Novel human prostate specific protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 98.6%; Score 1151; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.7e-278;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCAGCCCTGGCAGCGCGCAGCTGTCATGAAACGAATGTCTCTCGGCGCTCTG 60
DB 1 GCGCAGCCCTGGCAGCGCGCAGCTGTCATGAAACGAATGTCTCTCGGCGCTCTG 60
QY 61 GTGCATCCGACAGTGGTCTGTCTACGCCGACACTGTTTCCAGAACTCTACACCATCGGG 120
DB 61 GTGCATCCGACAGTGGTCTGTCTACGCCGACACTGTTTCCAGAACTCTACACCATCGGG 120
QY 121 CTGGGCTGACAGTCTTGAGGCGCCAGCAAGAGCCAGGAGCCAGATGTGTGAGGCGCAGC 180
DB 121 CTGGGCTGACAGTCTTGAGGCGCCAGCAAGAGCCAGGAGCCAGATGTGTGAGGCGCAGC 180
QY 121 CTGGGCTGACAGTCTTGAGGCGCCAGCAAGAGCCAGGAGCCAGATGTGTGAGGCGCAGC 180
DB 121 CTGGGCTGACAGTCTTGAGGCGCCAGCAAGAGCCAGGAGCCAGATGTGTGAGGCGCAGC 180
QY 181 CTCTCCGTACGCGACCCAGAGTACACAGACTCTTGCTCGCTAACGACCTCATGCTCATC 240
DB 181 CTCTCCGTACGCGACCCAGAGTACACAGACTCTTGCTCGCTAACGACCTCATGCTCATC 240
QY 241 AAGTTGACGAATCCGTGTCGAGTCTGACACACCATCCGAGCATCAGCATGCTTCGACAG 300
DB 241 AAGTTGACGAATCCGTGTCGAGTCTGACACACCATCCGAGCATCAGCATGCTTCGACAG 300
QY 301 TGCCCTACCGCGGGGGAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 301 TGCCCTACCGCGGGGGAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 361 ATGCCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420

| | | | |
|----|------|--|------|
| Db | 361 | ATGCTTACCGTGTGCTGCACTGCGCTGAACGTGTGCGTGTGTCTGAGANGTCTGACAGTAAG | 420 |
| QY | 421 | CTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCGCGGAGGCGCAAGACCAGAAAG | 480 |
| Db | 421 | CTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCGCGGAGGCGCAAGACCAGAAAG | 480 |
| QY | 481 | GACTCTCGCAACGGTGACTCTGGGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTT | 540 |
| Db | 481 | GACTCTCGCAACGGGTACTTGCAGGGGTACTTGCAGGGCCTT | 540 |
| QY | 541 | GTCGCTTTGCGGAAAAGCCCCGTGTGGCCCACTTGGCGGTGCCAGGTGTCTACACCAACCTC | 600 |
| Db | 541 | GTCGCTTTGCGGAAAAGCCCCGTGTGGCCCACTTGGCGGTGCCAGGTGTCTACACCAACCTC | 600 |
| QY | 601 | TGCAAAATTCAGTGTGATGAGAGAAAACCGTCCAGNCCAGTTAACTCTGGGGACTGGGA | 660 |
| Db | 601 | TGCAAAATTCAGTGTGATGAGAGAAAACCGTCCAGNCCAGTTAACTCTGGGGACTGGGA | 660 |
| QY | 661 | ACCCATGAAATTGACCCCAATACATCTCGGGAANGAATTCAGGAATATCTGTTCCCA | 720 |
| Db | 661 | ACCCATGAAATTGACCCCAATACATCTCGGGAANGAATTCAGGAATATCTGTTCCCA | 720 |
| QY | 721 | GCCCCCTCTCCCTCAGAGCCCAAGAGTCCAGGCCCCCAGCCCCCTCTCCCTCAACCAAGG | 780 |
| Db | 721 | GCCCCCTCTCCCTCAGAGCCCAAGAGTCCAGGCCCCCAGCCCCCTCTCCCTCAACCAAGG | 780 |
| QY | 781 | GTCACAGATCCCCAGCCCCCTCTCTCCCTCAGACCAGAGATCCAGACCCCCAGCCCCCTCNT | 840 |
| Db | 781 | GTCACAGATCCCCAGCCCCCTCTCTCCCTCAGACCAGAGATCCAGACCCCCAGCCCCCTCNT | 840 |
| QY | 841 | CCNTCAGACCCAGAGAGTCCAGCCCCCTCTCCNTCAGACCGCAGAGTCCAGACCCCCAGC | 900 |
| Db | 841 | CCNTCAGACCCAGAGAGTCCAGCCCCCTCTCCNTCAGACCGCAGAGTCCAGACCCCCAGC | 900 |
| QY | 901 | CCNTCNTCCGTCCAGACCCAGGGGTGCAGGCCCCCAACCCCTCTCNTCAGAGTCAGAGG | 960 |
| Db | 901 | CCNTCNTCCGTCCAGACCCAGGGGTGCAGGCCCCCAACCCCTCTCNTCAGAGTCAGAGG | 960 |
| QY | 961 | TCCAAGCCCCCAACCCCTCTGTTCCCCAGACCCAGAGGTNCAGGTCCAGCCCCCTCTCCC | 1020 |
| Db | 961 | TCCAAGCCCCCAACCCCTCTGTTCCCCAGACCCAGAGGTNCAGGTCCAGCCCCCTCTCCC | 1020 |
| QY | 1021 | TCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACACAGTGCCCCCCTGTGGCA | 1080 |
| Db | 1021 | TCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACACAGTGCCCCCCTGTGGCA | 1080 |
| QY | 1081 | NGTTGACCCCAACCTTACCAGTTGGTTTTTCATTTTGTGCCCTTTCCCTAGATCCAGAA | 1140 |
| Db | 1081 | NGTTGACCCCAACCTTACCAGTTGGTTTTTCATTTTGTGCCCTTTCCCTAGATCCAGAA | 1140 |
| QY | 1141 | ATAAAGTNTAAGAGAAAGCGCAAAAAA 1167 | |
| Db | 1141 | ATAAAGTNTAAGAGAAAGCGCAAAAAA 1167 | |

RESULT 2

V61252

ID V61252 standard; cDNA; 1167 BP.

AC V61252;

DT 06-JAN-1999 (first entry)

DE cDNA sequence of prostate tumour clone P703 splice variant DE13.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens.

PN W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI: 98-609886/51.

DR P-PSDB: W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
of a prostate tumour protein. The encoded immunogen, or the DNA itself
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

| | | | | |
|----------------------------|--------|---------------------|-----------|--------------|
| Query Match | 98.6% | Score 1151; | DB 1; | Length 1167; |
| Best Local Similarity | 100.0% | Pred. No. 2.7e-278; | | |
| Matches 1167; Conservative | 0; | Mismatches 0; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| QY | 1 | GCGAGCCCTGGCAGGCGGCACCTGGTCATGGA AAAA CCAATTGTCTGCTGGGCGTCTG | 60 |
| Db | 1 | GCGAGCCCTGGCAGGCGGCACCTGGTCATGGA AAAA CCAATTGTCTGCTGGGCGTCTG | 60 |
| QY | 61 | GTGCATCCGAGTGGGTGTGTCAGCCGCACACTGTTTCCAGA CTCTACACCATCGGG | 120 |
| Db | 61 | GTGCATCCGAGTGGGTGTGTCAGCCGCACACTGTTTCCAGA CTCTACACCATCGGG | 120 |
| QY | 121 | CTGGGCTTGACAGTCTTGAGGCCGACAAGAGCCAGGAGCCAGATGGTGAGGCCAGC | 180 |
| Db | 121 | CTGGGCTTGACAGTCTTGAGGCCGACAAGAGCCAGGAGCCAGATGGTGAGGCCAGC | 180 |
| QY | 181 | CTCTCCGTACGGCACCCAGATAACAAGACTCTGCTCGCTAACGACCTCATGCTCATC | 240 |
| Db | 181 | CTCTCCGTACGGCACCCAGATAACAAGACTCTGCTCGCTAACGACCTCATGCTCATC | 240 |
| QY | 241 | AAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTGGCAG | 300 |
| Db | 241 | AAGTTGAGCAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTGGCAG | 300 |
| QY | 301 | TGCCCTACCGGGGGAACTCTTGCCCTGNTCTGGCTGGGGTCTGCTGGCGAACGGCAGA | 360 |
| Db | 301 | TGCCCTACCGGGGGAACTCTTGCCCTGNTCTGGCTGGGGTCTGCTGGCGAACGGCAGA | 360 |
| QY | 361 | ATGCTACCGTGTCTGACACTGCGTGAACGTGTCGGTGTGTCTGAGANGTCTGCAGTAAG | 420 |
| Db | 361 | ATGCTACCGTGTCTGACACTGCGTGAACGTGTCGGTGTGTCTGAGANGTCTGCAGTAAG | 420 |
| QY | 421 | CTGTATGACCCGCTGTATACCA CCCAGCATGTTCTGCGCCGGGAGGCAAGACCAGAAG | 480 |
| Db | 421 | CTGTATGACCCGCTGTATACCA CCCAGCATGTTCTGCGCCGGGAGGCAAGACCAGAAG | 480 |
| QY | 481 | GACTCTGCAACGGTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGCAAGGGCTT | 540 |
| Db | 481 | GACTCTGCAACGGTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGCAAGGGCTT | 540 |
| QY | 541 | GTTGTCTTGGAAAAAGCCCCCGTGGCCCAACTTGCGGTGCCAGGTGTCTACACCAACCTC | 600 |
| Db | 541 | GTTGTCTTGGAAAAAGCCCCCGTGGCCCAACTTGCGGTGCCAGGTGTCTACACCAACCTC | 600 |
| QY | 601 | TGCAAA TCACTGAGTAGAGAAAACCGTCCAGNCCAGTTA CTCTGGGACTGGGA | 660 |
| Db | 601 | TGCAAA TCACTGAGTAGAGAAAACCGTCCAGNCCAGTTA CTCTGGGACTGGGA | 660 |
| QY | 661 | ACCCA TGAAT TGACCCCAAATFACATCTCGGAANGAATTCAGAA TATCTGTCCCA | 720 |
| Db | 661 | ACCCA TGAAT TGACCCCAAATFACATCTCGGAANGAATTCAGAA TATCTGTCCCA | 720 |
| QY | 721 | GCCCCTCTCCCTCAGGCCAGAGTCCAGGCCCCCAGCCCCCTCTCTCCA AACCAAGG | 780 |
| Db | 721 | GCCCCTCTCCCTCAGGCCAGAGTCCAGGCCCCCAGCCCCCTCTCTCCA AACCAAGG | 780 |
| QY | 781 | GTACGATCCCCAGCCCCCTCTCTCTCAGACCCAGAGTCCAGACCCCCAGCCCCCTCNT | 840 |
| Db | 781 | GTACGATCCCCAGCCCCCTCTCTCTCAGACCCAGAGTCCAGACCCCCAGCCCCCTCNT | 840 |


```
QY 841 CCNTCAGACCCAGAGTCCAGCCCCCTCTCCNTCAGACGAGGAGTCCAGACCCCCAGC 900
Db 841 CCNTCAGACCCAGAGTCCAGCCCCCTCTCCNTCAGACGAGGAGTCCAGACCCCCAGC 900
QY 901 CCNTCNTCCGTACAGCCAGGGGTGACAGGCCCCCAACCCCTCNCNTCAGAGTACAGAG 960
Db 901 CCNTCNTCCGTACAGCCAGGGGTGACAGGCCCCCAACCCCTCNCNTCAGAGTACAGAG 960
QY 961 TCCAAGCCCCCAACCCCTCGTTCGCCAGACCCAGAGGTNCAGGTCCAGCCCTCTCCC 1020
Db 961 TCCAAGCCCCCAACCCCTCGTTCGCCAGACCCAGAGGTNCAGGTCCAGCCCTCTCCC 1020
QY 1021 TCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACACAGTCCCCCTGTGGA 1080
Db 1021 TCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACACAGTCCCCCTGTGGA 1080
QY 1081 NGTTGACCCCAACCTTACCAGTTGTTTTCATTTTGTCCCTTTCCCTAGATCCAGAA 1140
Db 1081 NGTTGACCCCAACCTTACCAGTTGTTTTCATTTTGTCCCTTTCCCTAGATCCAGAA 1140
QY 1141 ATAAAGTNTAAGAGAGCGCAAAAAA 1167
Db 1141 ATAAAGTNTAAGAGAGCGCAAAAAA 1167

RESULT 3
ID V58644 standard; cDNA; 1248 BP.
AC V58644;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE1.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 217..696
FT /tag= a
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
DR P-PSDB; W69387.
PT Novel human prostate specific tumour protein and fragments - useful
PS for detecting and treating prostate cancers
PS Claim 1; Page 112; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;
```

```
Query Match 95.2%; Score 1111.2; DB 1; Length 1248;
Best Local Similarity 96.7%; Pred. No. 2.3e-268;
Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps 1;
```

```
QY 2 CGCAGCCCTGGCAGCGGCACTGTCTATGAAAAAGATTGTTCTCTCGGCGTCTG 61
Db 41 CGCAGCCCTGGCAGCGGCACTGTCTATGAAAAAGATTGTTCTCTCGGCGTCTG 100
QY 62 TGCATCCGAGTGGTGTCTGTCAGCCGCACTGTTCCAGAA-----CTCCT 109
Db 101 TGCATCCGAGTGGTGTCTGTCAGCCGCACTGTTCCAGAAAGTGCAGAGCTCCT 160
```

```
QY 110 ACACATCGGGCTGGGCTGACACATCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 169
Db 161 ACACATCGGGCTGGGCTGACACATCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 220
QY 170 TGGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACTTGTGCTGCTAACGACC 229
Db 221 TGGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACTTGTGCTGCTAACGACC 280
QY 230 TCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCGAGCATCAGCA 289
Db 281 TCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCGAGCATCAGCA 340
QY 290 TTGCTTGGAGTGGCCCTACCGGGGGAACTCTTGCCCTCCTNCTGGCTGGGCTGCTGG 349
Db 341 TTGCTTGGAGTGGCCCTACCGGGGGAACTCTTGCCCTCCTNCTGGCTGGGCTGCTGG 400
QY 350 CGAAGCGCAAAATGCCCTACCGGTGCTGACCTGCGTGAACGTGTCGGTGTCTGAGGANG 409
Db 401 CGAAGCGCAAAATGCCCTACCGGTGCTGACCTGCGTGAACGTGTCGGTGTCTGAGGANG 460
QY 410 TCTGAGTAAGCTCTATGACCCGCTGTACACCCAGCATGTCTGCGCCGGCGAGGGC 469
Db 461 TCTGAGTAAGCTCTATGACCCGCTGTACACCCAGCATGTCTGCGCCGGCGAGGGC 520
QY 470 AAGACCAAGAGACTCCTGCAACGGTGACTGTGGGGGCCCTGATCTGCAACGGGTACT 529
Db 521 AAGACCAAGAGACTCCTGCAACGGTGACTGTGGGGGCCCTGATCTGCAACGGGTACT 580
QY 530 TGCAGGGCTTGTGCTTTCGGAAGAGCCCGTGTGGCCAACCTGGCGTGCCAGGTGCT 589
Db 581 TGCAGGGCTTGTGCTTTCGGAAGAGCCCGTGTGGCCAACCTGGCGTGCCAGGTGCT 640
QY 590 ACACCAACCTCTGCAAAATCCTACTGAGTGATAGAGAAAACGTCACAGNCCAGTTAACTCT 649
Db 641 ACACCAACCTCTGCAAAATCCTACTGAGTGATAGAGAAAACGTCACAGNCCAGTTAACTCT 700
QY 650 GGGGACTGGGAACCCATGAAATGACCCCAATATACATCTCGGAANGAATTCAGGAAT 709
Db 701 GGGGACTGGGAACCCATGAAATGACCCCAATATACATCTCGGAANGAATTCAGGAAT 760
QY 710 ATCTGTTCCAGCCCTCTCTCCCTCAGGCCAGAGTCCAGAGGCCCTCTCTCC 769
Db 761 ATCTGTTCCAGCCCTCTCTCCCTCAGGCCAGAGTCCAGAGGCCCTCTCTCC 820
QY 770 TCAACCAAGGGTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCC 829
Db 821 TCAACCAAGGGTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCC 880
QY 830 CAGCCCTCTCCTCCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCA 889
Db 881 CAGCCCTCTCCTCCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCA 940
QY 890 GACCCCCAGCCCTCCTCCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCA 949
Db 941 GACCCCCAGCCCTCCTCCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGTCCA 1000
QY 950 AGAGTCAAGGTCCAAAGCCCAACCCCTCGTTCCTCAGACCCAGAGTNCAGGTCCAG 1009
Db 1001 AGAGTCAAGGTCCAAAGCCCAACCCCTCATATCCCAAGAGTNCAGGTCCAG 1060
QY 1010 CCCCTCTCTCTCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACAGAGTGGC 1069
Db 1061 CCCCTCTCTCTCAGACCCAGCGGTCCAATGCCACCTAGANTNTCCCTGTACAGAGTGGC 1120
QY 1070 CCCTTGTGGCANGTTGACCCCAACCTTACCAGTGTGTTTTCATTTTGTGCTTCC 1129
Db 1121 CCCTTGTGGCANGTTGACCCCAACCTTACCAGTGTGTTTTCATTTTGTGCTTCC 1180
QY 1130 TAGATCCAGAAATAAGTNTAAGAGAGCGCAAAAAA 1167
Db 1181 TAGATCCAGAAATAAGTNTAAGAGAGNGCAAAAAA 1218
```


RESULT 4
ID V61249 standard; cDNA; 1248 BP.
AC V61249;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE1.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 95.2%; Score 1111.2; DB 1; Length 1248;
Best Local Similarity 96.7%; Pred. No. 2.3e-268;
Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps 1;

QY 2 CGAGCCCTGGCAGGCGGCGACTGTCATGGAAGCAATGTCTGCTCGGCGCTCTGG 61
DB 41 CGAGCCCTGGCAGGCGGCGACTGTCATGGAAGCAATGTCTGCTCGGCGCTCTGG 100
QY 62 TGCATCCGCACTGGGTGCTGTACGCCGCACTGTTCAGAA-----CTCCT 109
DB 101 TGCATCCGCACTGGGTGCTGTACGCCGCACTGTTCAGAAAGTAGTGACAGCTCCT 160
QY 110 ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGG 169
DB 161 ACACCATCGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGG 220
QY 170 TGGAGGCCAGCCTCTCCGTACGCGCACAGAGTACAAACAGACTCTTGTCTCGTAACGACC 229
DB 221 TGGAGGCCAGCCTCTCCGTACGCGCACAGAGTACAAACAGACTCTTGTCTCGTAACGACC 280
QY 230 TCATGCTCATCAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCA 289
DB 281 TCATGCTCATCAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCA 340
QY 290 TTGCTTCGCACTGCCCTACCGCGGGGAACCTTGCCTCGTNTCTGGCTGGGGTCTGCTGG 349
DB 341 TTGCTTCGCACTGCCCTACCGCGGGGAACCTTGCCTCGTNTCTGGCTGGGGTCTGCTGG 400
QY 350 CGAACGCGAGAATGCCCTACCGTGTCTGCACTGCGTGAACGTTGCGGTGTCTGAGGANG 409
DB 401 CGAACGCGAGAATGCCCTACCGTGTCTGCACTGCGTGAACGTTGCGGTGTCTGAGGANG 460
QY 410 TCTGAGTAAGCTCTATGACCCGCTGTACACACCCAGACATGTCTGCGCGCGGAGGGC 469
DB 461 TCTGAGTAAGCTCTATGACCCGCTGTACACACCCAGACATGTCTGCGCGCGGAGGGC 520
QY 470 AAGACCAAGAGACTCTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT 529
DB 521 AAGACCAAGAGACTCTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACT 580
QY 530 TGCAGGGCCTTGTGCTTTCGAAAGCCCGGTGTGCGCAACTTGGCGTGCCAGGTGCT 589

DB 581 TGCAGGGCCTTGTGCTTTCGAAAGCCCGTGTGCGCAAGTTGGCGTGCCAGGTGCT 640
QY 590 ACACCAACTCTGCAAAATCACTGAGTGATAGAAACCGTCCAGNCCAGTTACTCT 649
DB 641 ACACCAACTCTGCAAAATCACTGAGTGATAGAAACCGTCCAGNCCAGTTACTCT 700
QY 650 GGGGACTGGGAACCCATGAATTGACCCCAATACATCTGCGGANGAATTCAGGAAT 709
DB 701 GGGGACTGGGAACCCATGAATTGACCCCAATACATCTGCGGANGAATTCAGGAAT 760
QY 710 ATCTGTCCAGCCCTCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCCAGCCCTCTCTCC 769
DB 761 ATCTGTCCAGCCCTCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCCAGCCCTCTCTCC 820
QY 770 TCAAAACCAAGGTACAGATCCCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCC 829
DB 821 TCAAAACCAAGGTACAGATCCCCAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCC 880
QY 830 CAGCCCT 889
DB 881 CAGCCCT 940
QY 890 GACCCCCAGCCCT 949
DB 941 GACCCCCAGCCCT 1000
QY 950 AGAGTCAGAGTCCAGACCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1009
DB 1001 AGAGTCAGAGTCCAGACCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1060
QY 1010 CCCCT 1069
DB 1061 CCCCT 1120
QY 1070 CCTTGTGGCANGTTGACCCCAACCTTACCAGTTGGTTTTCATTTTGTCCCTTCTCCC 1129
DB 1121 CCTTGTGGCANGTTGACCCCAACCTTACCAGTTGGTTTTCATTTTGTCCCTTCTCCC 1180
QY 1130 TAGATCCAGAAATAAGTTAAGAGAAGCGCAAAAAA 1167
DB 1181 TAGATCCAGAAATAAGTTAAGAGAAGNGCAAAAAA 1218

RESULT 5
ID V58645 standard; cDNA; 1265 BP.
AC V58645;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.

| | | | | | | |
|-------------|----------|---|--------|---------------|--------|--------------------------------------|
| SQ | Sequence | 1265 BP; | 256 A; | 432 C; | 321 G; | 245 T; |
| Query Match | | | | | | |
| | | 88.8%; | | Score 1036.2; | | DB 1; |
| | | Best Local Similarity | | 91.0%; | | Pred. No. 1.2e-249; |
| | | Matches 1137; | | Conservative | | 0; Mismatches 29; Indels 83; Gaps 1; |
| OY | 2 | CGCAGCCCTGGCAGCGGCACACTGTCATGGAAGAATGTCTGCTCGGGCTCTGG | 61 | | | |
| DB | 14 | CGCAGCCCTGGCAGCGGCACACTGTCATGGAAGAATGTCTGCTCGGGCTCTGG | 73 | | | |
| OY | 62 | TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAACTCCTACACCATCGGC | 121 | | | |
| DB | 74 | TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAACTCCTACACCATCGGC | 133 | | | |
| OY | 122 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 181 | | | |
| DB | 134 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 193 | | | |
| OY | 182 | TCTCCGTACGGCACCAGAGTACAACAGACTCTGCTCGCTAACGACCTCATGCTATCA | 241 | | | |
| DB | 194 | TCTCCGTACGGCACCAGAGTACAACAGACCTTGCTCTAACGACCTCATGCTATCA | 253 | | | |
| OY | 242 | AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATGCTTCGAGT | 301 | | | |
| DB | 254 | AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATGCTTCGAGT | 313 | | | |
| OY | 302 | GCCCTACCGCGGGGAACCTCTGCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAC----- | 355 | | | |
| DB | 314 | GCCCTACCGCGGGGAACCTCTGCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAC | 373 | | | |
| OY | 355 | ----- | 355 | ----- | ----- | ----- |
| DB | 374 | TCACGGGTGTGTGTCTGCCCTCTCAAGAGGTCCTCTGCCAGTCGCGGGGGCTGACCC | 433 | | | |
| OY | 355 | -----GGCAGAATGCCCTACCGTGTGCTGACCTGCGTGAACGTGTGGTGT | 398 | | | |
| DB | 434 | AGAGCTGTGCTGCCAGGCGAGAATGCTTACCGTGTGCTGAGTGCCTGAACGTGTGGTGT | 493 | | | |
| OY | 399 | GTTCTGAGGANGTCTGAGTAACTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC | 458 | | | |
| DB | 494 | GTTCTGAGGANGTCTGAGTAACTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC | 553 | | | |
| OY | 459 | CGCGGAGGGAAGCAGGAAGACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTG | 518 | | | |
| DB | 554 | CGCGGAGGGAAGCAGGAAGACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTG | 613 | | | |
| OY | 519 | CAACGGGTACTGTGAGGGCCCTGTGTCTTTCGGAAGAGCCCGTGTGGCCAACTGGCGT | 578 | | | |
| DB | 614 | CAACGGGTACTGTGAGGGCCCTGTGTCTTTCGGAAGAGCCCGTGTGGCCAACTGGCGT | 673 | | | |
| OY | 579 | GCCAGGTGTCTACCAACCTCTGCAAAATCACTGAGTGATAGAGAAAACCGTCCAGNC | 638 | | | |
| DB | 674 | GCCAGGTGTCTACCAACCTCTGCAAAATCACTGAGTGATAGAGAAAACCGTCCAGNC | 733 | | | |
| OY | 639 | CAGTTAACTCTGGGACTGGGAACCCATGAATTGACCCCAAAATACATCTGCGGAAG | 698 | | | |
| DB | 734 | CAGTTAACTCTGGGACTGGGAACCCATGAATTGACCCCAAAATACATCTGCGGAAG | 793 | | | |
| OY | 699 | AATTAGGAATATCTGTCCAGCCCTCTCTCCCTCAGGCCAGAGATCCAGGCCCCAG | 758 | | | |
| DB | 794 | AATTAGGAATATCTGTCCAGCCCTCTCTCCCTCAGGCCAGAGATCCAGGCCCCAG | 853 | | | |
| OY | 759 | CCCCCTCTCCCTCAACCAAGGTACAGATCCCCAGCCCTCTCTCTCAGACCCAGAG | 818 | | | |
| DB | 854 | CCCCCTCTCCCTCAACCAAGGTACAGATCCCCAGCCCTCTCTCTCAGACCCAGAG | 913 | | | |
| OY | 819 | TCCAGACCCCCAGCCCT | 878 | | | |
| DB | 914 | TCCAGACCCCCAGCCCT | 973 | | | |
| OY | 879 | GCAGGAGTCCAGACCCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 938 | | | |

| | | | |
|---|---|--|------|
| Db | 974 | CCAGAGTCCAGACCCCCAGCCCTCTCTCTCTCAGACCCAGCGGGTTGAGCCCCCAACC | 1033 |
| OY | 939 | CCTCTCTCTCAGAGTTCAGAGGTTCCAAAGCCCCCAACCCCTGTTCCCGCAGACCGAGGT | 998 |
| Db | 1034 | CCTCTCTCTCAGAGTTCAGAGGTTCCAAAGCCCCCAACCCCTGTTCCCGCAGACCGAGGT | 1093 |
| OY | 999 | NCAGTCCAGCCCCCTCTCTCTCTCAGACCCAGCGGTCCAATGCGACCTAGANTNTCCCTG | 1058 |
| Db | 1094 | NNAGTCCAGCCCCCTCTCTCTCTCAGACCCAGNGTCCCAATGCGACCTAGATTTCCCTG | 1153 |
| OY | 1059 | TACACAGTGGCCCCCTGTGTGGCANGTTGACCCCAACCTTACAGTGTGTTTTCATTTTGG | 1118 |
| Db | 1154 | NACACAGTGGCCCCCTGTGTGGNANGTTGACCCCAACCTTACAGTGTGTTTTCATTTTNG | 1213 |
| OY | 1119 | TCCCTTTCCCTAGATTCAGAAATAAGTNTAAGAGAAGCGCAAAAAA 1167 | |
| Db | 1214 | TCCCTTTCCCTAGATTCAGAAATAAGTNTAAGAGANGCAAAAAA 1262 | |
| RESULT 6 | | | |
| ID | V61250 | standard; cDNA; 1265 BP. | |
| AC | V61250; | | |
| DT | 06-JAN-1999 | (first entry) | |
| DE | cDNA sequence of prostate tumour clone P703 splice variant DE2. | | |
| OS | Prostate; cancer; tumour; vaccine; immunogen; clone; ss. | | |
| KW | Homo sapiens. | | |
| PN | W09837093-A2. | | |
| PD | 27-AUG-1998. | | |
| PF | 25-FEB-1998; U03492. | | |
| PR | 09-FEB-1998; US-020956. | | |
| PR | 25-FEB-1997; US-806099. | | |
| PR | 01-AUG-1997; US-904804. | | |
| PA | (CORI-) CORIXA CORP. | | |
| PI | Dillon DC, Xu J; | | |
| DR | WPI; 98-609886/51. | | |
| PT | Polypeptides comprising immunogenic portions of prostate proteins - | | |
| PT | used in a vaccine for the treatment of prostate cancer | | |
| PS | Claim 3; Page 105-106; 130pp; English. | | |
| CC | The present sequence is a new DNA which encodes an immunogenic portion | | |
| CC | of a prostate tumour protein. The encoded immunogen, or the DNA itself, | | |
| CC | can be used as a vaccine for the treatment of prostate cancer. The DNA | | |
| CC | was identified by analysis of a subtracted cDNA library obtained by | | |
| CC | subtracting a prostate tumour cDNA expression library with a normal | | |
| CC | tissue cDNA library. | | |
| SQ | Sequence | 1265 BP; 256 A; 432 C; 321 G; 245 T; | |
| Query Match 88.8%; Score 1036.2; DB 1; Length 1265; | | | |
| Best Local Similarity 91.0%; Pred. No. 1.2e-249; | | | |
| Matches 1137; Conservative 0; Mismatches 29; Indels 83; Gaps 1; | | | |
| OY | 2 | CGCAGCCCTGGCAGCGGCACACTGTCATGGAAGAAAGAAATGTTCTGCTCGGGCTCTGG | 61 |
| Db | 14 | CGCAGCCCTGGCAGCGGCACACTGTCATGGAAGAAAGAAATGTTCTGCTCGGGCTCTGG | 73 |
| OY | 62 | TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAACTCCTACACCATCGGC | 121 |
| Db | 74 | TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAACTCCTACACCATCGGC | 133 |
| OY | 122 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 181 |
| Db | 134 | TGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 193 |
| OY | 182 | TCTCCGTACGGCACCAGAGTACAACAGACTCTGCTCGCTAACGACCTCATGCTATCA | 241 |
| Db | 194 | TCTCCGTACGGCACCAGAGTACAACAGACCTTGCTCGCTAACGACCTCATGCTATCA | 253 |
| OY | 242 | AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGATTGCTTCGAGT | 301 |
| Db | 254 | AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGATTGCTTCGAGT | 313 |
| OY | 302 | GCCCTACCGCGGGGAACCTCTGCCTCGTNTCTGGCTGGGGTCTGCTGGCGAAC----- | 355 |

| | | | |
|----|------|--|------|
| OY | 470 | AAGACCAGAAGGACTCCTGCACCGGTGACTCTGGGGGGCCCTGATCTGCACCGGGTACT | 529 |
| Db | 501 | AAGACCAGAAGGACTCCTGCACCGGTGACTCTGGGGGGCCCTGATCTGCACCGGGTACT | 560 |
| OY | 530 | TGCAGGGCCCTGTGTCTTTGCGAAAAAGCCCCGTGTGGCCAATTGGCGTGCCAGGTGTC | 589 |
| Db | 561 | TGCAGGGCCCTGTGTCTTTGCGAAAAAGCCCCGTGTGGCCAATTGGCGTGCCAGGTGTC | 620 |
| OY | 590 | ACACCAACCTCTGCAAATTCAGTAGTGGATAGAGAAAACCGT-CCAGNCCAGTTAACTC | 648 |
| Db | 621 | ACACCAACCTCTGCAAATTCAGTAGTGGATAGAGAAAACCGTACCAGGCCAGTTAACTC | 680 |
| OY | 649 | TGGGGACTGGGAACCCATGAAATTGACCCCCCAAATACATCCTGCGAANGAATTACGAA | 708 |
| Db | 681 | TGGGGACTGGGAACCCATGAAATTGACCCCCCAAATACATCCTGCGAANGAATTACGAA | 740 |
| OY | 709 | TATCTGTCCAGGCCCTCTCCCTCAGGCCCAGAGTCCAGGGCCCCCAGCCCCTCTCC | 768 |
| Db | 741 | TATCTGTCCAGGCCCTCTCCCTCAGGCCCYCAGGAGTCCAGGGCCCCCAGCCCCTCTCC | 800 |
| OY | 769 | CTCAAACCAAGGGTACAGATCCCCAGCCCCCTCTCCTCAGACCAGAGTCCAGACCC | 828 |
| Db | 801 | CTCAAACCAAGGGTACAGATCCCCAGCCCCCTCTCCTCAGACCAGAGTCCAGACCC | 860 |
| OY | 829 | CCAGCCCCCTCNTCCNTCAGACCAGAGTCCAGCCCCCTCTCNTCAGACGACGAGATCC | 888 |
| Db | 861 | CCAGCCCCCTCTCCTCCTCAGACCCAGAGTCCAGCCCCCTCTCCTCAGACCCAGAGTCC | 920 |
| OY | 889 | AGAC----- | 893 |
| Db | 921 | AGACCCCCCAGCCCCCTCTCCTCAGACCAGGGGTCAGCCTCTCTCCTCAGACCCA | 980 |
| OY | 893 | ----- | 893 |
| Db | 981 | GGAGTCCAGACCCCCCAGCCCCCTCTCCTCAGACCAGAGTCCAGCCCCCTCTCTCTC | 1040 |
| OY | 893 | -----CCCCCAGCCCNMTCGCTCAGACCCAGGGGTGACGGCCCC | 934 |
| Db | 1041 | AGACCCAGAGTCCAGATCCCCCAGCCCCCTCTCTCCTCAGACCCAGGGGTCCAGGCCCC | 1100 |
| OY | 935 | AACCCCTCNTCCNTCAGAGTCCAGAGTCCCAAGCCCCCAACCCCTCGTTCOCAGACCCAG | 994 |
| Db | 1101 | AACCCCTCTCTCTCAGACTCAGAGGTCCAAAGCCCCCAACCCCTCTTCCCCAGACCCAG | 1160 |
| OY | 995 | AGGTNCAGGTCCAGCCCCCTCTCTCCTCAGACCCAGCGGTCCATGCCCACCTAGANTMC | 1054 |
| Db | 1161 | AGGTNCAGGTACACGCCCTCTCTCTCCTCAGACCCAGCGGTCCATGCCCACCTATACTTC | 1220 |
| OY | 1055 | CCTGTACACAGTGGCCCCCTGTGGCANGTGAACCAACCTTACCAGTGTGTTTTCATTT | 1114 |
| Db | 1221 | CCTGTACANATTTGCCNCCTGTGTGGCACGTTGACCCCAACCTTACCAGTGTGTTTTCATTT | 1280 |
| OY | 1115 | TTTGTCCCTTTCCCTTATATCCAGAATAAAGTNTAAGAGAAGCGCAAAAAA 1167 | |
| Db | 1281 | TTTGTCCCTTTCCCTTATATCCAGAATAAAGTNTAAGRAGRAGSGCCAAAAAA 1333 | |

| RESULT | 8 |
|--------|--|
| V37495 | |
| ID | V37495 standard; DNA; 871 BP. |
| AC | V37495; |
| DT | 07-SEP-1998 (first entry) |
| DE | Human prostate-specific kallikrein (HPSK) encoding DNA. |
| KW | prostate-specific kallikrein; HPSK; prostate carcinoma; human; |
| KW | benign prostate hyperplasia; diagnosis; drug screening; PSK; ss. |
| OS | Homo sapiens. |
| PH | Key |
| FT | CDS |
| FT | location/Qualifiers |
| FT | 31..777 |
| FT | /*tag= a |
| FT | /transl_except= (pos:367..369, aa:Xaa) |
| FT | /transl_except= (pos:412..414, aa:Xaa) |
| FT | /transl_except= (pos:424..426, aa:Xaa) |
| FT | /product= "HPSK protein" |

FT /note= "Xaa - unknown"
 PN W09820117-A1.
 PD 14-MAY-1998.
 PF 31-OCT-1997; U20051.
 PR 05-NOV-1996; US-744026.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Goll SK;
 DR WPI; 98-286933/25.
 DR P-PSDB; W60592.
 PT New isolated prostate-specific kallikrein - used to develop products
 PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
 PT hyperplasia
 PS Claim 5; Fig 1A-C; 68pp; English.
 CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
 CC containing an expression vector comprising the HPSK nucleic acid sequence
 CC can be used to produce the protein recombinantly. The HPSK products can
 CC be used for the diagnosis of conditions or diseases associated with
 CC expression of HPSK such as prostate carcinoma and benign prostate
 CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
 CC modulate its activity can be used for the preparation of treatment of
 CC such conditions or diseases. The products can also be used for detection
 CC and drug screening, especially for the detection of prostate-specific
 CC kallikrein (PSK).
 SO Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

| Query Match | 61.8%; | Score 721.4; | DB 1; | Length 871; |
|-----------------------|---|---------------------|----------------|-------------|
| Best Local Similarity | 98.2%; | Pred. No. 3.6e-171; | | |
| Matches 725; | Conservative | 0; | Mismatches 13; | Indels 0; |
| | | | | Gaps |
| QY 2 | CGCAGCCCTGGCAGCGCGGCACCTGGTCAATGAAAAAGAAATGTTCTGCTCGGGCGTCTCG | 61 | | |
| | | | | |
| Db 134 | CCCAGCCCTGGCAGCGCGGCACCTGGTCAATGAAAAAGAAATGTTCTGCTCGGGCGTCTCG | 193 | | |
| QY 62 | TGCATCCGCAATGGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCATCGGGC | 121 | | |
| | | | | |
| Db 194 | TGCATCCGCAATGGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCATCGGGC | 253 | | |
| QY 122 | TGGGCGTGCACAGTCTTGAGGCCGACCAAGCCAGGAGCCAGATGGTGGAGGCCAGCC | 181 | | |
| | | | | |
| Db 254 | TGGGCGTGCACAGTCTTGAGGCCGACCAAGCCAGGAGCCAGATGGTGGAGGCCAGCC | 313 | | |
| QY 182 | TCTCCGTACGGCACCAGAGTACACAGACTCTTGCTCGCTAACGACCTCATGTCTATCA | 241 | | |
| | | | | |
| Db 314 | TCTCCGTACGGCACCAGAGTACACAGACTCTTGCTCGCTAACGACCTCATGTCTATCA | 373 | | |
| QY 242 | AGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGACATCAGCATTTGCTTGGCAGT | 301 | | |
| | | | | |
| Db 374 | AGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGATTCAGCATTTGCTTGGCAGT | 433 | | |
| QY 302 | GCCCTACCCGGGGGAACCTTGTGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 361 | | |
| | | | | |
| Db 434 | GCCCTACCCGGGGGAACCTTGTGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGCGCAGAA | 493 | | |
| QY 362 | TGCCTACCGTGTCTGCACTGCGTGAACGTGTCGGTGTCTGAGGANGTCTGCAGTAAGC | 421 | | |
| | | | | |
| Db 494 | TGCCTACCGTGTCTGCACTGCGTGAACGTGTCGGTGTCTGAGGANGTCTGCAGTAAGC | 553 | | |
| QY 422 | TCTATGACCCGCTGTACACACCCAGCAATGTTCTGCGCCGCGGAGGCAAGACAGAGAG | 481 | | |
| | | | | |
| Db 554 | TCTATGACCCGCTGTACACACCCAGCAATGTTCTGCGCCGCGGAGGCAAGACAGAGAG | 613 | | |
| QY 482 | ACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGACAGGGCCTTG | 541 | | |
| | | | | |
| Db 614 | ACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGACAGGGCCTTG | 673 | | |
| QY 542 | TGCTTTTCGGAAGAGCCCGTGTGGCCAACTTGGCGTGCCAGGTCTACACCAACCTCT | 601 | | |
| | | | | |
| Db 674 | TGCTTTTCGGAAGAGCCCGTGTGGCCAACTTGGCGTGCCAGGTCTACACCAACCTCT | 733 | | |
| QY 602 | GCAAAATTCAGTGTGATAGAGAAAACGTCAGNCCAGTTAACTCTGGGGACTGGGAA | 661 | | |
| | | | | |
| Db 734 | GCAAAATTCAGTGTGATAGAGAAAACGTCAGNCCAGTTAACTCTGGGGACTGGGAA | 793 | | |

QY 662 CCCATGAATGACCCCAATACATCTCGCGAANGAATTCAGGAATATCTGTCCAG 721
DB 794 CCCATGAATGACCCCAATACATCTCGCGAANGAATTCAGGAATATCTGTCCAG 853
QY 722 CCCCTCTCTCCCTCAGGCC 739
DB 854 CCCCTCTCTCCCTCAGGCC 871

RESULT 9

V58648
ID V58648 standard; cDNA; 1119 BP.
AC V58648;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT CDS /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69389.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 116-117; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 30.2%; Score 352.2; DB 1; Length 1119;
Best Local Similarity 96.3%; Pred. No. 4.8e-79;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CGCAGCCCTGGCAGCGGCACATGTCATGGAAGAATGTTCTGCTCGGCGTCTGG 61
DB 8 CGCAGCCCTGGCAGCGGCACATGTCATGGAAGAATGTTCTGCTCGGCGTCTGG 67
QY 62 TGCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCAGAACTCTTACACCATCGGGC 121
DB 68 TGCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCAGAACTCTTACACCATCGGGC 127
QY 122 TGGGCTTGACAGTCTTGAGGCGGACCAAGAGAGCCAGAGCCAGATGTTGAGGCGCAGCC 181
DB 128 TGGGCTTGACAGTCTTGAGGCGGACCAAGAGAGCCAGATGTTGAGGCGCAGCC 187
QY 182 TCTCCGTACGGCAGCCAGAGTACAAAGAGACTCTTGCCTAACGACCTCATGCTCATCA 241
DB 188 TCTCCGTACGGCAGCCAGAGTACAAAGAGACTCTTGCCTAACGACCTCATGCTCATCA 247
QY 242 AGTTGACGAATCCGTGCTCGAGTCTGACACCATCCGAGAGCATCAGCATTTGCTTCGAGT 301
DB 248 AGTTGACGAATCCGTGCTCGAGTCTGACACCATCCGAGAGCATCAGCATTTGCTTCGAGT 307
QY 302 GCCCTACCGCGGGAGACTCTTGCCTCTGTTCTGCTGGGCTCTGCTGGCGAAGCGCAGAA 361

DB 308 GCCCTACCGCGGGAGACTCTTGCCTCTGTTCTGCTGGGCTGCTGCGGAGACGATGCTG 367
QY 362 TGCCCTACCGTGTG 375
DB 368 TGATGGCATCCAG 381

RESULT 10

V61253
ID V61253 standard; cDNA; 1119 BP.
AC V61253;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT CDS /*tag= a
PN WO9837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 108-109; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 30.2%; Score 352.2; DB 1; Length 1119;
Best Local Similarity 96.3%; Pred. No. 4.8e-79;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CGCAGCCCTGGCAGCGGCACATGTCATGGAAGAATGTTCTGCTCGGCGTCTGG 61
DB 8 CGCAGCCCTGGCAGCGGCACATGTCATGGAAGAATGTTCTGCTCGGCGTCTGG 67
QY 62 TGCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCAGAACTCTTACACCATCGGGC 121
DB 68 TGCATCCGAGTGGGTGCTGTACAGCCGACACTGTTCAGAACTCTTACACCATCGGGC 127
QY 122 TGGGCTTGACAGTCTTGAGGCGGACCAAGAGAGCCAGAGCCAGATGTTGAGGCGCAGCC 181
DB 128 TGGGCTTGACAGTCTTGAGGCGGACCAAGAGAGCCAGATGTTGAGGCGCAGCC 187
QY 182 TCTCCGTACGGCAGCCAGAGTACAAAGAGACTCTTGCCTAACGACCTCATGCTCATCA 241
DB 188 TCTCCGTACGGCAGCCAGAGTACAAAGAGACTCTTGCCTAACGACCTCATGCTCATCA 247
QY 242 AGTTGACGAATCCGTGCTCGAGTCTGACACCATCCGAGAGCATCAGCATTTGCTTCGAGT 301
DB 248 AGTTGACGAATCCGTGCTCGAGTCTGACACCATCCGAGAGCATCAGCATTTGCTTCGAGT 307
QY 302 GCCCTACCGCGGGAGACTCTTGCCTCTGTTCTGCTGGGCTCTGCTGGCGAAGCGCAGAA 361
DB 308 GCCCTACCGCGGGAGACTCTTGCCTCTGTTCTGCTGGGCTCTGCTGGCGAAGCGCAGAA 367
QY 362 TGCCCTACCGTGTG 375
DB 368 TGATGGCATCCAG 381

RESULT 11

X41114
ID X41114 standard; cDNA; 402 BP.

AC X41114;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:58.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR P-PSDB; Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12514 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 28.4%; Score 331.4; DB 1; Length 402;
Best Local Similarity 96.2%; Pred. No. 5.5e-74;
Matches 354; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

OY 2 CGCAGCCCTGGCAGCGGCGACTGTCATGGAACGAATTGTTCTGCGGCGCTCTGG 61
DB 29 CGCAGCCCTGGCAGCGGCGACTGTCATGGAACGAATTGTTCTGCGGCGCTCTGG 88
OY 62 TGCATCCGAGTGGTGTCTGACGCCGACACTGTTTCCAGAA-----CTCCT 109
DB 89 TGCATCCGAGTGGTGTCTGACGCCGACACTGTTTCCAGAAAGTAGTACAGAGCTCCT 148
OY 110 ACACCATCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 169
DB 149 ACACCATCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 208
OY 170 TGGAGGCCAGCCTCTCCGTCAGGACCCAGAGTACACAGACTCTTGCTGCTAACGACC 229
DB 209 TGGAGGCCAGCCTCTCCGTCAGGACCCAGAGTACACAGACTCTTGCTGCTAACGACC 268
OY 230 TCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTACACCATCCGAGCATCAGCA 289
DB 269 TCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTACACCATCCGAGCATCAGCA 328
OY 290 TTGCTTCGAGTGGCCCTACCGCGGGGAACTCTTGCTCGTNTCTGGCTGGGCTCTGCTGG 349
DB 329 TTGCTTCGAGTGGCCCTACCGCGGGGAACTCTTGCTCGTNTCTGGCTGGGCTCTGCTGG 388
OY 350 CGAACGGC 357
DB 389 CGAACGGC 396

RESULT 12
ID V58646
AC V58646;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone D6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 25.1%; Score 292.8; DB 1; Length 1459;
Best Local Similarity 80.4%; Pred. No. 3.4e-64;
Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

OY 92 ACTGTTCCAGAACCTCTACACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 151
DB 25 AGTGAAGTGCAGAGCTCTACACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 84
OY 152 AGCCAGGAGCCAGATGTTGAGAGCCAGCTCTCCGTACGGCACCAGATACACAGAC 211
DB 85 AGCCAGGAGCCAGATGTTGAGAGCCAGCTCTCCGTACGGCACCAGATACACAGAC 144
OY 212 TCTTCTGCTAACGACCTCATGCTCATAGTTGAGCAATCCGTGTCGAGTCTGACA 271
DB 145 CTTGCTGCTAACGACCTCATGCTCATAGTTGAGCAATCCGTGTCGAGTCTGACA 204
OY 272 CCATCCGAGCATCAATGCTTCCGAGTCCCTACCGCGGGAATCTTGCCCTGTTT 331
DB 205 CCATCCGAGCATCAATGCTTCCGAGTCCCTACCGCGGGAATCTTGCCCTGTTT 264
OY 332 CTGGCTGGGCTGCTGGCGAAC----- 355
DB 265 CTGGCTGGGCTGCTGGCGAACGGGTGAGCTCAGGGGTGTGTCTGCCCTCTCAAGGA 324
OY 355 -----GGCAGATGCTTAC 368
DB 325 GGTCTCTGCCAGTCCGGGGGCTGACCCAGAGCTCTGCGTCCAGGACAGATGCTTAC 384
OY 369 CGTGTGCACTGCGTGAACGTGTGGTGTCTGAGGANGTCTGCAAGTAACTTATGA 428
DB 385 CGTGTGCACTGCGTGAACGTGTGGTGTCTGAGGANGTCTGCAAGTAACTTATGA 444
OY 429 CCCGCTGTACACCCAGCATGTTCTGGCGCGGAGGAGGCAAGCAAGAGACTCTG 488
DB 445 CCCGCTGTACACCCAGCATGTTCTGGCGCGGAGGAGGCAAGCAAGAGACTCTG 504
OY 489 CAACG 493
DB 444 CAACG 493

PS Claim 3, page 53-54; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 19.68; | Score 228.8; | DB 1; | Length 234; |
| Best Local Similarity | 98.38; | Pred. No. 1.9e-48; | | |
| Matches 230; Conservative | 0; | Mismatches 4; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 203 | ACACAGAGCTCTTGGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCC | 262 |
| | | | |
| Db | 1 | ACACAGAGACCTTGGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCC | 60 |
| QY | 263 | AGCTGACACCATCCGGAGCATCAGATTGCTTCGCAGTGCCTACCGCGGGAAGCTTT | 322 |
| | | | |
| Db | 61 | AGCTGACACCATCCGGAGCATCAGATTGCTTCGCAGTGCCTACCGCGGGAAGCTTT | 120 |
| QY | 323 | GCCTCGTTCGGCTGGGGTCTGCTGGCGAAGCGCAGAATGCTTACCGTGCCTGACGCG | 382 |
| | | | |
| Db | 121 | GCCTCGTTCGGCTGGGGTCTGCTGGCGAAGCGCAGAATGCTTACCGTGCCTGACGCG | 180 |
| QY | 383 | TGAACGTGTCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTTATGACCCGCTGT | 436 |
| | | | |
| Db | 181 | TGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCAGTAAGCTTATGACCCGCTGT | 234 |

Search completed: September 25, 1999, 09:57:15
Job time: 8086 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:23 ; Search time 1811.29 seconds
(without alignments)
1270.888 Million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167

Sequence: 1 GCGCAGCCCTGCGAGCGGC.....NTAAGAGAAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs; 986266752 residues

Database :

EST: *
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 555.6 | 47.6 | 777 | 48 | AI557281 | AI557281 PT2.1_15_ |
| 2 | 403.2 | 34.6 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 3 | 393 | 33.7 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 4 | 348 | 29.8 | 404 | 34 | AA503963 | AA503963 nh39a01.s |
| 5 | 286.6 | 24.6 | 576 | 34 | AA533140 | AA533140 nj46h06.s |
| 6 | 221 | 18.9 | 722 | 48 | AI557025 | AI557025 PT2.1_10_ |
| 7 | 217.2 | 18.6 | 259 | 35 | AA565161 | AA565161 nk52c07.s |
| 8 | 210.8 | 18.1 | 229 | 35 | AA552779 | AA552779 nk57a10.s |
| 9 | 203 | 17.4 | 218 | 50 | AI674521 | AI674521 wc39g02.x |
| 10 | 201.2 | 17.2 | 241 | 36 | AA603977 | AA603977 no47a09.s |
| 11 | 174 | 14.9 | 188 | 50 | AI675815 | AI675815 wb97b12.x |
| 12 | 164 | 14.1 | 216 | 32 | AA336074 | AA336074 EST40886 |
| 13 | 161.8 | 13.9 | 191 | 35 | AA595489 | AA595489 no34h04.s |
| 14 | 132 | 11.3 | 457 | 45 | AI385433 | AI385433 m185b05.y |
| 15 | 118.2 | 10.1 | 498 | 26 | W73168 | W73168 zd55e11.rl |
| 16 | 116 | 9.9 | 505 | 45 | AI391329 | AI391329 mb71h07.y |
| 17 | 114 | 9.8 | 496 | 39 | AA864127 | AA864127 vn42g07.r |
| 18 | 112.6 | 9.6 | 507 | 35 | AA583052 | AA583052 nn80g09.s |
| 19 | 106.4 | 9.1 | 517 | 31 | AA293027 | AA293027 zt54a12.r |
| 20 | 100.2 | 8.6 | 590 | 33 | AA411252 | AA411252 zt33b03.r |
| 21 | 99.2 | 8.5 | 759 | 45 | AI326340 | AI326340 m161e08.x |
| 22 | 99 | 8.5 | 321 | 31 | AA302930 | AA302930 EST113097 |
| 23 | 98.4 | 8.4 | 686 | 46 | AI415008 | AI415008 mb71h07.x |
| 24 | 98 | 8.4 | 586 | 26 | W73140 | W73140 zd55e11.sl |
| 25 | 97.4 | 8.3 | 626 | 36 | AA643312 | AA643312 nr59b08.s |
| 26 | 96 | 8.2 | 333 | 49 | AI636241 | AI636241 tz93c03.x |
| 27 | 92.6 | 7.9 | 356 | 39 | AA838788 | AA838788 a187g08.s |
| 28 | 91.4 | 7.8 | 212 | 28 | C16642 | C16642 C16642 C1on |
| 29 | 90.2 | 7.7 | 420 | 43 | AI226226 | AI226226 ue88f06.y |
| 30 | 90 | 7.7 | 585 | 35 | C23111 | C23111 C23111 Japa |
| 31 | 90 | 7.7 | 504 | 49 | AI620091 | AI620091 ty48d10.x |
| 32 | 89.8 | 7.7 | 480 | 39 | AA846771 | AA846771 a141f01.s |
| 33 | 89.8 | 7.7 | 367 | 41 | AI002101 | AI002101 ot38d03.s |
| 34 | 89.6 | 7.7 | 234 | 33 | AA452459 | AA452459 zx29g09.r |
| 35 | 89 | 7.6 | 400 | 31 | AA293231 | AA293231 zt26g09.r |
| 36 | 89 | 7.6 | 539 | 34 | AA477689 | AA477689 zu44a12.r |
| 37 | 87.8 | 7.5 | 467 | 28 | AA073833 | AA073833 mj99h09.r |
| 38 | 87.8 | 7.5 | 460 | 38 | AA791893 | AA791893 vs54h07.r |
| 39 | 87.4 | 7.5 | 519 | 33 | AA401397 | AA401397 zu68b01.s |
| 40 | 87.4 | 7.5 | 583 | 40 | AA921373 | AA921373 ak56d12.s |
| 41 | 87.4 | 7.5 | 493 | 45 | AI324874 | AI324874 m185b05.x |
| 42 | 87.2 | 7.5 | 541 | 39 | AA844955 | AA844955 ak61d09.s |
| 43 | 87.2 | 7.5 | 324 | 43 | AI177474 | AI177474 EST221106 |
| 44 | 87.2 | 7.5 | 599 | 43 | AI237604 | AI237604 EST234166 |
| 45 | 87.2 | 7.5 | 457 | 46 | AA998593 | AA998593 vt-R-CO-1 |

ALIGNMENTS

RESULT 1
AI557281
LOCUS AI557281 777 bp MRNA
DEFINITION PT2.1_15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557281
NID g4489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 777)
AUTHORS Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Mathew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11: 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda Zp II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."

BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 47.6%; Score 555.6; DB 48; Length 777;
Best Local Similarity 91.8%; Pred. No. 1.3e-128;
Matches 637; Conservative 0; Mismatches 46; Indels 11; Gaps 5;

QY 2 CGCAGCCCTGGCAGCGCGGCTGTCATGGAAAAAGCAATTGTTCTGCTCGGGGCTCCTGG 61
DB 21 CGCAGCCCTGGCAGCGCGGCTGTCATGGAAAAAGCAATTGTTCTGCTCGGGGCTCCTGG 80

QY 62 TGCATCCGCAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAACTCCTACACCATGGGC 121
DB 81 TGCATCCGCAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAACTCCTACACCATGGGC 140

QY 122 TGGGCTGACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGCC 181
DB 141 TGGGCTGACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGCC 200

QY 182 TCTCCGTACGGGACCCAGAGTACAAAGACTCTTGCTCGCTAACGACCTCATGCTATCA 241
DB 201 TCTCCGTACGGGACCCAGAGTACAAAGACTCTTGCTCGCTAACGACCTCATGCTATCA 260

QY 242 AGTGGACGAATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCATTTGCTTCGAGT 301
DB 261 AGTGGACGAATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCATTTGCTTCGAGT 320

QY 302 GCCCTACCGCGGGGAACCTCTGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGGGCAGAA 361
DB 321 GCCCTACCGCGGGGAACCTCTGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAGGGCAGAA 380

QY 362 TGCCTACCGTGTGCTGACCTGCGAAGCTGTGGTGTCTGAGGANGTCTGAGTAAGC 421
DB 381 TG-CTACCGTGTGCTGACCTGCGAAGCTGTGGTGTCTGAGGANGTCTGAGTAAGC 439

QY 422 TCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGCAAGACAGAAG 481
DB 440 TCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGCAAGACAGAAG 499

QY 482 ACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 541
DB 500 ACTCTGCAACGGTGACTCTGGGGGG-CCCTGATNTGNAACGGGNCCTTGCAGGGCCTTG 558

QY 542 TGTCTTCGAAAGCCCGGTGTGGCCACTTGGCGGTGCCAGGTGTCTACACCACTCT 601
DB 559 TGTCTTCGAAAGCCCGGTGTGG--CAAGTTGGCGGTGCAGGTGTACACAAC-----CT 611

QY 602 GCAATTCAGTGTGATAGAGAAACCGCTCCAGNCCAGTTACTCTGGGACTGGGA 661
DB 612 CTGAATTTACTGAGTGTAGANAAAACGGTCAGG--CAGTTACTNTGGGNACTGGNA 669

QY 662 CCCATGAATTGACCCCAATATCATCTCGCGGA 695
DB 670 NCCATTAATTGNCCCAAAATANATCTGGGGA 703

RESULT 2
AI686689/c
LOCUS
DEFINITION
AI686689 569 bp mRNA EST 27-MAY-1999
tu35g11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253092 3'
similar to TR:Q92046 Q92046 PREPROTRYPsin PRECURSOR ; mRNA
sequence.
ACCESSION
NID AI686689
VERSION AI686689.1 GI:4897983
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 569)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 444.

FEATURES
source
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7/3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 150 c 170 g 124 t
ORIGIN

Query Match 34.6%; Score 403.2; DB 50; Length 569;
Best Local Similarity 98.8%; Pred. No. 1.1e-90;
Matches 405; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 151 GAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCCACCCAGAGTACAACAGA 210
|||||
Db 569 GAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCCACCCAGAGTACAACAGA 510
QY 211 CTCTGCTCGCTAACGACCTCATGTCTATCAAGTTGGACGAATCCGTGTCGAGTGTAC 270
|||||
Db 509 CCCTGCTCGCTAACGACCTCATGTCTATCAAGTTGGACGAATCCGTGTCGAGTGTAC 450
QY 271 ACCATCCGAGACATCAGCATTTGCTTCGAGTGCCTACCCGGGGAACCTTTGCCCTGTN 330
|||||
Db 449 ACCATCCGAGACATCAGCATTTGCTTCGAGTGCCTACCCGGGGAACCTTTGCCCTGT 390
QY 331 TCTGCTGGGCTCTGCTGGCGAAGCGGAGAATGCCCTACCGTGTGCACTGCGTGAACGTG 390
|||||
Db 389 TCTGCTGGGCTCTGCTGGCGAAGCGGAGAATGCCCTACCGTGTGCACTGCGTGAACGTG 330
QY 391 TCGGTGCTGTGAGGANGTCTGCAGTACCTCTATGACCCGCTGTACACCCAGCATG 450
|||||
Db 329 TCGGTGCTGTGAGGANGTCTGCAGTACCTCTATGACCCGCTGTACACCCAGCATG 270
QY 451 TTCTGCGCCGCGGAGGGCAAGACCAGAAGACTCTCTGCAACCGGTGACTTGGGGGGCCC 510
|||||
Db 269 TTTTGGCGCCGCGGAGGGCAAGACCAGAAGACTCTCTGCAACCGGTGACTTGGGGGGCCC 210
QY 511 CTGATCTGCAACGGGTACTGTGAGGGCCTGTGTCTTTGGAAGAGCCCC 560
|||||
Db 209 CTGATCTGCAACGGGTACTGTGAGGGCCTGTGTCTTTGGAAGAGCCCC 160

RESULT 3
AA551449 415 bp mRNA EST 05-SEP-1997
LOCUS nJ55e05.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996416
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo;
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 640 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 412.
FEATURES
Location/Qualifiers
1. 415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_lib="NCI_CGAP_Pr9"
/sex="male"

/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."

BASE COUNT 78 a 125 c 129 g 83 t
ORIGIN
Query Match 33.7%; Score 393; DB 35; Length 415;
Best Local Similarity 98.1%; Pred. No. 3.3e-88;
Matches 407; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 137 TTGAGCCGACCAAGAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCACC 196
|||||
Db 2 TCGACGCCGACCAAGAGCCAGGAGCCAGACGGGTGAGGCCAGCCTCTCCGTACGGCACC 61
QY 197 CAGAGTACAACAGACTCTGCTGCTTAACGACCTCATGCTCATCAAGTTGAGCAATCCG 256
|||||
Db 62 CAGAGTACAACAGACCCCTGCTGCTTAACGACCTCATGCTCATCAAGTTGAGCAATCCG 121
QY 257 TGTCCGAGTCTGACACCATCCGAGCATCAGCATTTGCTTCGACGTGCCCTACCGGGGA 316
|||||
Db 122 TGTCCGAGTCTGACACCATCCGAGCATCAGCATTTGCTTCGACGTGCCCTACCGGGGA 181
QY 317 ACTCTTGCTCGTNTCTGCGTGGGCTCTGCTGGCGAAGCGGACAGATGCCCTACGCTGC 376
|||||
Db 182 ACTCTTGCTCGTNTCTGCGTGGGCTCTGCTGGCGAAGCGGACAGATG-CTACCGTCTGC 240
QY 377 ACTGCGTGAACGTGTGCGTGTCTGAGGANGTCTGCAGTAAAGCTCTATGACCCGCTGT 436
|||||
Db 241 AGTGGGTGAACGTGTGCGTGTCTGAGGANGTCTGCAGTAAAGCTCTATGACCCGCTGT 300
QY 437 ACCACCCGACATGTTCTGCGCGCGGAGGGCAAGACAGAGGACTCTGCAACGGTG 496
|||||
Db 301 ACCACCCGACATGTTCTGCGCGCGGAGGGCAAGACAGAGGACTCTGCAACGGTG 360
QY 497 ACTCTGGGGGCCCTGATCTGCAACGGGTACTGTGACGGCCTTGTGCTTTGCG 551
|||||
Db 361 ACTCTGGGGGCCCTGATCTGCAACGGGTACTGTGACGGCCTTGTGCTTTGCG 415

RESULT 4
AA503963 404 bp mRNA EST 20-AUG-1997
LOCUS nh39a01.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954696
DEFINITION similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; contains MSRI.b2 MSRI repetitive element ; mRNA sequence.
ACCESSION AA503963
NID 92238930
VERSION AA503963.1 GI:2238930
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo;
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 8, 1995 this sequence version replaced gi:801235.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

| | | |
|-----------|--|------------|
| VERSION | AI557025.1 | GI:4489388 |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. | |
| AUTHORS | 1 (bases 1 to 722) Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L. | |
| TITLE | Prostate Cancer Expression Profiling by cDNA Sequencing Analysis | |
| JOURNAL | Unpublished (1999) | |
| COMMENT | On May 18, 1998 this sequence version replaced gi:3138511. | |

CONTACT: Guyang Matthew Huang
 Leroy Hood
 University of Washington
 Department of Molecular Biotechnology, Box 357730, University of
 Washington, Seattle, WA 98195
 Tel: 5106280100
 Fax: 5106280108
 Email: huanggm@yahoo.com.
LOCATION/QUALIFIERS
 1. .722

```

/note="Organ: Prostate; Vector: pBluescript; Directional
cdna library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
```

| BASE COUNT | ORIGIN |
|------------|-----------------------------|
| 146 a | 156 c 195 g 176 t 49 others |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 18.98; | Score 221; | DB 48; | Length 722; |
| Best Local Similarity | 94.88; | Pred. No. 2.9e-45; | | |
| Matches 235; Conservative | 0; | Mismatches 12; | Indels 1; | Gaps 1; |

| | | | | | | | | | | |
|----|-----|-----|------------------|--------------------|------------------|--------------------|----------------------|------------|----------------------|-----|
| OY | 103 | AAC | TCC | TAC | CACC | ATCGGGCTGGG | CCCTGC | CACAGTCTTG | AGGCCCAAGAGCCAGGGAGC | 162 |
| | | I | I | I | I | I | I | I | I | I |
| Dd | 115 | AGC | TTCTA | CACCATCGGGCTGGG | CCCTGC | CACAGTCTTG | AGGCCCAAGAGCCAGGGAGC | 174 | | |
| OY | 163 | CAG | ATGTGGAGG | CCAGCCCTCTCCGTA | CGGCACCCAGAGTACA | CAGACTCTTGCTGCCT | 222 | | | |
| | | I | I | I | I | I | I | I | I | I |
| Dd | 175 | CAG | ATGTGGAGG | CCAGCCCTCTCCGTA | CGGCNA | CCAGNGTACA | CAGACCCTTGCTGCCT | 234 | | |
| OY | 223 | AAC | GACCTCATGCTCATCA | AGTTGAGCAATCCG | TGTCCGAGTCTGA | CACCATCCGGAGC | 282 | | | |
| | | I | I | I | I | I | I | I | I | I |
| Dd | 235 | AAC | GACCTCATGCTCATNA | AGTTGAGCAATCCG | TGTCCGAGTCTGA | NACCATCCGGAGC | 294 | | | |
| OY | 283 | ATC | AGCATTCGCTTCG | CAGTGCCCTACGCGGGG | GAAC | TCTTGCC | -TCGINTCTGCGTGGG | 341 | | |
| | | I | I | I | I | I | I | I | I | I |
| Dd | 295 | ATC | AGCATTCGCTTCG | NAAGTGCCCTACCGTGGG | GAAC | TCTTNCCITTCGTTTCTG | GCGTGGG | 354 | | |
| OY | 342 | TCT | GCTGG | 349 | | | | | | |
| | | I | I | I | I | I | | | | |
| Dd | 355 | TCT | GCTTG | 362 | | | | | | |

| RESULT | 7 | | | | | | | | |
|------------|---|--------------|--------------|---|-------------|--|--|--|--|
| LOCUS | AA565161/c | | | | | | | | |
| DEFINITION | AA565161 | 259 bp | mRNA | EST | 08-SEP-1997 | | | | |
| | nk52c07.s1 | NCI_CGAP_Pr7 | Homo sapiens | cdna clone IMAGE:1017132 | | | | | |
| | | | | similar to contains MSRI.b3 MSRI MSRI repetitive element ; , mRNA | | | | | |
| | | | | sequence. | | | | | |
| ACCESSION | AA565161 | | | | | | | | |
| NID | q2336800 | | | | | | | | |
| VERSION | AA565161.1 | GI:2336800 | | | | | | | |
| KEYWORDS | EST. | | | | | | | | |
| SOURCE | human. | | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | | | | | | |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eutheria: Primates: Catarrhini; Homiinae; Homo.
1 (bases 1 to 259)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1397569.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 458 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 258.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .259 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q21"
/clone="IMAGE:1017132"
/clone_1lb="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from prostate intraepithelial neoplasia (high-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."

```

| BASE COUNT | 56 a | 41 c | 106 g | 56 t |
|---------------------------|------|--------|--------------------|-------------------|
| ORIGIN | | | | |
| Query Match | | 18.6%; | Score 217.2; | DB 35; |
| Best Local Similarity | | 91.8%; | Pred. No. 1.7e-44; | Length 259; |
| Matches 225; Conservative | | 0; | Mismatches 20; | Indels 0; Gaps 0; |

| | | | | | | | | | | |
|----|------|--------------|------------|----------|----------|------------|----------|------------|-----------|------|
| QY | 915 | ACCCAGGGGGTG | CAGGCCCCCA | ACCCCTC | NTCCNTC | CAGAGT | CAGAGGTC | CAAGCCCCCA | AC | 974 |
| | | | | | | | | | | |
| Db | 245 | AACCAAGGGT | ACAGATCCCC | AGCCCCCT | CTCTTC | CAGACC | CAGGAGTC | CAAGCCCCCA | AC | 186 |
| QY | 975 | CCCTCGTCCC | CAGACCAG | AGGTNC | AGGTCC | AGCCCCCT | CTCTC | CAGACC | CCAGCGGT | 1034 |
| | | | | | | | | | | |
| Db | 185 | CCCTCCTTCCC | CAGACCAG | AGGTCC | AGGTAC | CAAGCCCCCT | CTCTC | CAGACC | CCAGCGGT | 126 |
| QY | 1035 | CCAATGCCAC | CTAGANTT | TCCTGT | ACACAG | TGCCCCCT | TGTGGC | ANGTTG | ACCCAACCT | 1094 |
| | | | | | | | | | | |
| Db | 125 | CCAATGCCAC | CTAGACTT | CCCTGT | ACACAT | TGCCCCCT | TGTGGC | AGTTG | ACCCAACCT | 66 |
| QY | 1095 | TACCAAGT | TGTTTTTC | ATTTTTT | TGTCCCTT | TCCCTTAG | ATCCAGA | ATAAAGT | NTAAGAG | 1154 |
| | | | | | | | | | | |
| Db | 65 | TACCAAGT | TGTTTTTC | ATTTTTT | TGTCCCTT | TCCCTTAG | ATCCAGA | ATAAAGT | CTAAGAG | 6 |
| QY | 1155 | AAGCG | 1159 | | | | | | | |
| | | | | | | | | | | |
| Db | 5 | AAGCG | 1 | | | | | | | |

| RESULT | 8 | | | | |
|------------|------------|--------|--|-----|-------------|
| AA552779/c | | | | | |
| LOCUS | | | | | |
| DEFINITION | AA552779 | 229 bp | mRNA | EST | 08-SEP-1997 |
| | nk57a10.s1 | | NCL CGAP_pr7 Homo sapiens cDNA clone IMAGE:1017594 | | |
| | similar to | | contains MSRI.b3 MSRI repetitive element ; | | |
| | sequence. | | | | |

| | |
|-----------|--|
| ACCESSION | AA552779 |
| NID | 92323033 |
| VERSION | AA552779.1 |
| KEYWORDS | GI:2323033 |
| SOURCE | EST. |
| ORGANISM | human. |
| REFERENCE | Homo sapiens |
| AUTHORS | Eukaryota; Metazoa; Chordata; Granlata; Vertebrata; Mammalia; |
| TITLE | Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| JOURNAL | 1 (bases 1 to 229) |
| COMMENT | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), |
| | Tumor Gene Index |
| | Unpublished (1997) |
| | On Jan 25, 1995 this sequence version replaced gi:637760. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nlh.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 419 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 215.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .229 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1017594"
/clone_lib="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
prostate intraepithelial neoplasia (high-grade), cDNA
made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp."

```

| | | | | |
|------------|------|------|------|------|
| BASE COUNT | 49 a | 38 c | 94 g | 48 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 18.1%; | Score 210.8; | DB 35; | Length 229; |
| Best Local Similarity | 93.9%; | Pred. No. 6.5e-43; | | |
| Matches 215; Conservative | 0; | Mismatches 14; | Indels 0; | Gaps 0; |

[illegible]

| | | | | |
|------------|------------|---------------|--------------|-----------------------------|
| RESULT | 9 | | | |
| AI674521/c | | | | |
| LOCUS | AI674521 | 218 bp | mRNA | EST |
| DEFINITION | wc39q02.x1 | NCI_CGAP_Pr28 | Homo sapiens | CDNA clone IMAGE:2321042 3' |

| | |
|-----------|---|
| ACCESSION | similar to contans MSRL.t2 MSRL repetitive element ; , mRNA sequence. |
| NID | AI674521 |
| VERSION | 94875001 |
| KEYWORDS | AI674521.1 GI:4875001 |
| SOURCE | EST. |
| | human. |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 218)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ccgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Mar 16, 1998 this sequence version replaced gi:2961915.

JOURNAL Unpublished (1997)
COMMENT On Mar 16, 1998 this sequence version replaced gi:2961915

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.-G.E. Consortium/DNL at:
www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .218 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2321042"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Ronaldo."

```

| | | | | |
|------------|------|------|------|------|
| BASE COUNT | 48 a | 36 c | 85 g | 49 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 17.48; | Score 203; | DB 50; | length 218; |
| Best Local Similarity | 94.98; | Pred. No. 5.6e-41; | | |
| Matches 206; Conservative | 0; | Mismatches 11; | Indels 0; | Gaps 0; |

| | | | |
|----|------|--|------|
| QY | 944 | TCCTCAGAGTCAGAGGTCCAAAGCCCCCAACCCCTCGTTCCCCAGACCAGAGGTCAGG | 1003 |
| | | | |
| Db | 217 | TCCCTCAGACTCAGAGGTCCAAGCCCCCAACCCCTCCTTCCCCAGACCAGAGGTCAGG | 158 |
| QY | 1004 | TCCAGCCCCCTCCTCCTCCTCAGACCCAGCGGTCCCAATGCCACTAGANTNTCCCTGTACAC | 1063 |
| | | | |
| Db | 157 | TCCAGCCCCCTCCTCCTCCTCAGACCCAGCGGTCCCAATGCCACTAGACTCTCCCTGTACAC | 98 |
| QY | 1064 | AGTCCCCCTTGTGGCANGTTGACCCCAACCTTACCAGTGTGTTTTCATTTTGTCCCT | 1123 |
| | | | |
| Db | 97 | AGTCCCCCTTGTGGCACGTTGACCCCAACCTTACCAGTGTGTTTTCATTTTGTCCCT | 38 |
| QY | 1124 | TTCCTCAGATCCAGAAATAAAGTNTAAGAGAAGCGC | 1160 |
| | | | |
| Db | 37 | TTCCTCAGATCCAGAAATAAAGTCTAAGAAAAGCGC 1 | |

QY 1124 TTCCCCCTAGATCCAGAAATAAGTNTAAGAGAAAGCGC 1160
||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 37 TTCCCAAGATCCAGAAATAAGTCTAAGAAAAAGCGC 1


```
RESULT 10
AA603977/c 241 bp mRNA EST 28-OCT-1997
LOCUS no47a09.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1103800 3'
DEFINITION similar to contains MSRL.b3 MSRL repetitive element ;, mRNA
sequence.
ACCESSION AA603977
NID 92444547
VERSION AA603977.1 GI:2444547
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 241)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395285.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2084 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 231.
Location/Qualifiers
1. 241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1103800"
/clone_1lb="NCI_CGAP_Pr23"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
GAATTCGGCAGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."
BASE COUNT 46 a 37 c 78 g 80 t
ORIGIN

Query Match 17.2%; Score 201.2; DB 36; Length 241;
Best Local Similarity 96.2%; Pred. No. 1.6e-40;
Matches 203; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
RESULT 11
A1675815/c 188 bp mRNA EST 19-MAY-1999
LOCUS wb97b12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2313599 3',
DEFINITION mRNA sequence.
ACCESSION A1675815
NID 94876295
VERSION A1675815.1 GI:4876295
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133540.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2313599"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 43 a 31 c 61 g 53 t
ORIGIN

Query Match 14.9%; Score 174; DB 50; Length 188;
Best Local Similarity 97.2%; Pred. No. 8.9e-34;
Matches 174; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```


RESULT 12
AA336074 216 bp mRNA EST 21-APR-1997
LOCUS
DEFINITION EST40886 Endometrial tumor Homo sapiens cDNA 5' end similar to
similar to kallikrein family, mRNA sequence.
ACCESSION AA336074
NID 91988560
VERSION AA336074.1 GI:1988560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 216)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudex,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
JOURNAL based upon 83 million nucleotides of cDNA sequence
MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT 96026280
On Apr 14, 1993 this sequence version replaced gi:692773.
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
FEATURES
source Location/Qualifiers
1..216
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):138014"
/db_xref="taxon:9606"
/clone_lib="Endometrial tumor"
/sex="female"
/dev_stage="adult"
/note="Organ: endometrium; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 45 a 65 c 63 g 38 t 5 others
ORIGIN
Query Match 14.1%; Score 164; DB 32; Length 216;
Best Local Similarity 97.0%; Pred. No. 2.9e-31;
Matches 164; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 CGCAGCCCTGGCAGCGGCACTGTCATGGAAGAATGTTCTGCTCGGGCGTCTCG 61
|||||
Db 48 CGCAGCCCTGGCAGCGGCACTGTCATGGAAGAATGTTCTGCTCGGGCGTCTCG 107
QY 62 TGCATCCGCACTGGGTGCTGTACGCCGACACTGTTCCAGAACTCCTACACCATCGGGC 121
|||||

Db 108 TGCATCCGCACTGGGTGCTGTNAGCCCGCACACTGTTCCAGAACTCCTACACCATCGGGC 167
QY 122 TGGGCTTCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCT 170
|||||
Db 168 TGGGCTTCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCT 216
RESULT 13
AA595489 191 bp mRNA EST 18-SEP-1997
LOCUS AA595489/c
DEFINITION no34h04.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1102615 3',
mRNA sequence.
ACCESSION AA595489
NID 92410839
VERSION AA595489.1 GI:2410839
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 191)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 14, 1993 this sequence version replaced gi:692625.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html
FEATURES
source Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1..191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1102615"
/clone_lib="NCI_CGAP_Pr23"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
GAATTGGCAGCAG 3' 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."
BASE COUNT 43 a 28 c 57 g 63 t
ORIGIN
Query Match 13.9%; Score 161.8; DB 35; Length 191;
Best Local Similarity 95.9%; Pred. No. 9.8e-31;
Matches 163; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 998 TNCAGGTCCAGCCCTCTCTCCCTACAGACCCAGCGGTCCAAATGCCACTAGANTTCCT 1057
|||||
Db 191 TCCAGGTACACAGCCCTCTCTCTTACAGCCAGCGGTCCAAATGCCACTAGACTCTCCT 132
QY 1058 GTACACAGTGCCCTTGTGGCANGTTGACCAACCTTACAGTTGTTTTCATTTT 1117
|||||
Db 131 GTACACAGTGCCCTTGTGGCAGCTTGACCAACCTTACAGTTGTTTTCATTTT 72
QY 1118 GTCCCTTCCCTAGATCCAGAAATAAGTNTAAGAGAGCGCAAAAAA 1167
|||||
Db 71 GTCCCTTCCCTAGATCCAGAAATAAGTCTAAGAGAGCGCAAAAAA 22

| | |
|------------|---|
| RESULT | 14 |
| LOCUS | AI385433 |
| DEFINITION | AI385433 457 bp mRNA EST 27-JAN-1999 m185b05.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:473361 5' similar to gb:M22612 TRYPSINOGEN I PRECURSOR (HUMAN); gb:X04574 Mouse mRNA for preprotrypsin (MOUSE);, mRNA sequence. |
| ACCESSION | AI385433 |
| NID | q4198896 |
| VERSION | AI385433.1 GI:4198896 |
| KEYWORDS | EST. |
| SOURCE | house mouse. Mus musculus |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| REFERENCE | 1 (bases 1 to 457) |
| AUTHORS | Marrar,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelsing,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. |
| TITLE | The WashU-NCI Mouse EST Project 1999 |
| JOURNAL | Unpublished (1999) |
| COMMENT | On May 1, 1998 this sequence version replaced q1:3105815. |

Contact: Marra M/Mashu-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:284105
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Putative full length read
 vector to vector length is 813
 Seq primer: -40RP from G1bco
 High quality sequence stop: 413.
 Location/Qualifiers
 1..457

```

FEATURES
    source
        1. .457
            Location/Qualifiers
                organism="Mus musculus"
                db_xref="taxon:10090"
                map="5 q31.3-33.2; 16"
                clone="IMAGE:473361"
                clone_1lb="Soares mouse p3NMf19.5"
                dev_stage="19.5 dpc total fetus"
                lab_host="DH10B (ampicillin resistant)"
                note="Vector: pT7T3D (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTT 3'],
                    double-stranded cDNA was size selected, ligated to Eco RI
                    adapters (Pharmacia), digested with Not I and cloned into
                    the Not I and Eco RI sites of a modified pT7T3 vector
                    (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. RNA was kindly provided by
                    Dr. Minoru Ko (Wayne State University)."
```

| | | | | |
|---------------------------|-------|--------------------|-----------|-------------|
| Query Match | 11.3% | Score 132; | DB 45; | Length 457; |
| Best Local Similarity | 58.5% | Pred. No. 3.6e-23; | | |
| Matches 269; Conservative | 0; | Mismatches 182; | Indels 9; | Gaps 2; |

QY 143 CCGACCAAGAGCCAGGAGCCAGATGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGT 202

| | | | |
|----|-----|--|-----|
| Db | 1 | CTGTCATGAGTCTGGGCGACGAGATGTTCCAGGGAATCAATTCATCCCCACCCCGGTT | 60 |
| QY | 203 | ACAACAGACTCTTGGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCG | 262 |
| Db | 61 | ACTCCACCCCTGGCCACTCCAAATGACCTCATGCTCATCAAAATGAAACAGAAAAATCCGTG | 120 |
| QY | 263 | AGTCTGACACCAATCCGGAGCATCAGCATTGCTTCGCAGTGGCCCTACCGGGGAACCTTT | 322 |
| Db | 121 | ACTCTCACTCAAGTGAAGCCCGTCGAAATTGCTTGTGACTGTGCCACCGAGGGACTAGT | 180 |
| QY | 323 | GCCTCGTCTGGGCTGGGGTCTGCTGGCGAACGGC-----AGATGGCTACCGTGTGC | 376 |
| Db | 181 | GCATGGTGTCTGGCTGGGGGACACGACGACGACGACCATTAATCTCCCGAAAGTCTTCC | 240 |
| QY | 377 | ACTGCGTGAACGTGTGGGTGTGCTGAGGANGTCTGCAGTAAAGCTTATGACCCCGTGT | 436 |
| Db | 241 | AGTGCCCTGAATATTACTGTGCTCAGTGAAGGAGAGGTGTAAAACTCCTACCCAGACAGA | 300 |
| QY | 437 | ACCACCCCAAGCATGTTCTGCGCCGCGGAGGGCAAGACAGAAAGACTCCTGCAACGGTG | 496 |
| Db | 301 | TAGACAAGACCATGTTCTGCGCAGGTGATG---AAGAGGCGAGGAGACTCCTGCCAGGGTG | 357 |
| QY | 497 | ACTCTGGGGGCCCCGTGATCTGCAACAGGGTACTTGACAGGGCCTTGTCTTTCGAAAG | 556 |
| Db | 358 | ATTCCGGAGGTCTGTGCTGCTGCAATGGCAAGTTACAGGGCCTTGTCTGGGGTGATTT | 417 |
| QY | 557 | CCCCGTGTGGCCAATTGGCGTGCCAGGTGTTACACCAA | 596 |
| Db | 418 | TCCCCTGTGCTCAGCGGACAGACAGACCAAGGTGTTACACCAA | 457 |

RESULT 15
W73168
LOCUS
DEFINITION
w73168 498 bp mRNA EST 16-OCT-1996
zd55ell.r1 Soares_fetal_heart_NbH19W Homo sapiens cDNA clone
IMAGE:344588.5' similar to SW:TRYZ_CANFA P06872 TRYPSINOGEN,
ANIONIC PRECURSOR ;, mRNA sequence.
W73168
ACCESSION
NID g1383322
VERSION w73168.1 GI:1383322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 498)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 5, 1995 this sequence version replaced g1:798187.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LINL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 761 Std Error: 0.00
 Seq primer: ETPRimer
 High quality sequence stop: 455.
 Location/Qualifiers
 1..498
 /organism="Homo sapiens"
 /db_xref="GDB:1269963"
 /db_xref="taxon:9606"
 /map="4"
 /clone="IMAGE:344588"


```
/clone_lib="Soares_fetal_heart_NbHL19W"  
/sex="unknown"  
/dev_stage="19 weeks"  
/lab_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGGCGCCGACATCTTTTCTTTTCTTTT  
3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbHL19W."
```

BASE COUNT 116 a 150 c 128 g 104 t
ORIGIN

Query Match 10.1%; Score 118.2; DB 26; Length 498;
Best Local Similarity 61.0%; Pred. No. 1e-19;
Matches 227; Conservative 0; Mismatches 136; Indels 9; Gaps 2;

```
QY 275 TCCGGAGCATCAGCATTTGCTTCGACGTGCCCTACCGCGGGAACCTTGCCCTGCTTCTG 334  
  ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25 TCAGACCCATCAACGTCCTCTCTCATTTGCCCTCTGCTGGGACAAAGTGTGTTCTG 84  
  
QY 335 GCTGGGTTCTGCTGGCGAACGGCAGATG-----CCTACCGTGTCTGCACTGGGTGACG 388  
  ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 85 GCTGGGGGACAAACCAAGAGAGCCCCCAAGTGCACCTTCCCTAAGTCTCCAGTGTGAATA 144  
  
QY 389 TGTGGTGTGTCTTGAGANGTCTGACAGTAAGCTCTATGACCCGCTGTACCACCCAGCA 448  
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 145 TCAGCGTGTAGTACGAAAAAGGTGCGAGGATGCTTACCCGAGACAGATAGATGACACCA 204  
  
QY 449 TGTCTGCGCGCGCGGAGGAGGCAAGACAGAGGACTCCTGCAACGGTGACTCTGGGGGGC 508  
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 205 TGTCTGTGCGCGGTGACA--AAGCAGGTAGAGACTCCTGCCAGGGTGATTCTGGGGGGC 261  
  
QY 509 CCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGCTTTTCGAAAAAGCCCCGTGGCC 568  
  ||| ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||  
Db 262 CTGTGTTGTGCAATGGCTCCCTGCAAGGACTCGTGTCTGGGGAAGATTACCTTGTGCC 321  
  
QY 569 AACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAAATTCACCTGAGTAGAGAAAA 628  
  ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||  
Db 322 GGCCCAACAGACGGGTGTCTACAGCAACCTCTGCAAGTTCACCAAGTGCATCCAGGAAA 381  
  
QY 629 CCGTCCAGNCCA 640  
  ||| ||||| |||  
Db 382 CCATCCAGGCCA 393
```

Search completed: September 25, 1999, 12:05:28
Job time: 8132 sec

OM of: US-09-030-606-177 to: A_Geneseq_36:* out_format : pfs
Date: Sep 25, 1999 11:36 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:
-MODEL-frame+n2p.model -DEV-xlp
-Q/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29804/app_query.fasta.1
-DB-A_Geneseq_36-QFMT-fastan -SUFFIX-rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blousum62
-TRANS-human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-177
Query length: 1119
Database: A_Geneseq_36:*
Database sequences: 188963
Database length: 23686106
Search time (sec): 185.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|---------------------|------|--------|---------|---------|-----|----------------------------------|
| A_Geneseq_36:W69389 | + | 854.00 | 1499.98 | 2.7e-76 | 164 | Prostate tumour specific gene c |
| A_Geneseq_36:W71873 | + | 854.00 | 1499.98 | 2.7e-76 | 164 | Protein encoded by prostate tum |
| A_Geneseq_36:W59129 | + | 621.00 | 1083.44 | 3.1e-53 | 232 | Homo sapiens Tub Interactor (h |
| A_Geneseq_36:W60592 | + | 610.00 | 1063.24 | 3.8e-52 | 248 | Human prostate-specific kallikr |
| A_Geneseq_36:W69388 | + | 569.00 | 992.63 | 4.0e-48 | 205 | Prostate tumour specific gene c |
| A_Geneseq_36:W71872 | + | 569.00 | 992.63 | 4.0e-48 | 205 | Protein encoded by prostate tum |
| A_Geneseq_36:W69387 | + | 328.00 | 568.31 | 2.2e-24 | 159 | Prostate tumour specific gene c |
| A_Geneseq_36:W12881 | + | 318.00 | 559.95 | 1.5e-23 | 66 | Human 5' EST secreted protein SE |
| A_Geneseq_36:W10694 | + | 296.00 | 506.38 | 3.8e-21 | 260 | Human recombinant neuropsin, us |
| A_Geneseq_36:W12393 | + | 296.00 | 506.38 | 3.8e-21 | 260 | Mouse neuropsin protein. Nucle |
| A_Geneseq_36:W94493 | + | 292.50 | 499.85 | 8.5e-21 | 268 | Human kallikrein. Human kallikr |
| A_Geneseq_36:R67888 | + | 274.50 | 468.57 | 5.0e-19 | 253 | Human stratum corneum chymotryp |
| A_Geneseq_36:W05383 | + | 274.50 | 468.57 | 5.0e-19 | 253 | Human amyloid precursor protein |
| A_Geneseq_36:W71005 | + | 272.50 | 464.66 | 7.9e-19 | 262 | Human prostate-associated kallik |
| A_Geneseq_36:R84671 | + | 272.00 | 464.84 | 8.6e-19 | 237 | Mature kallikrein hK3. New isol |
| A_Geneseq_36:W45398 | + | 272.00 | 464.84 | 8.6e-19 | 237 | Prostate-specific antigen prote |
| A_Geneseq_36:P70568 | + | 271.50 | 462.88 | 1.0e-18 | 262 | Human kallikrein-like substance |
| A_Geneseq_36:W56086 | + | 270.00 | 461.29 | 1.3e-18 | 237 | Human prostate specific antigen |
| A_Geneseq_36:W83213 | + | 270.00 | 461.29 | 1.3e-18 | 237 | Prostate-specific antigen prote |
| A_Geneseq_36:W96187 | + | 270.00 | 461.29 | 1.3e-18 | 237 | Human prostate specific antigen |
| A_Geneseq_36:W13649 | + | 270.00 | 460.27 | 1.4e-18 | 261 | Human prostatic specific antigen |
| A_Geneseq_36:P92314 | + | 268.50 | 458.28 | 1.9e-18 | 245 | Human recombinant kallikrein gen |
| A_Geneseq_36:P95121 | + | 268.50 | 457.57 | 2.0e-18 | 262 | kallikrein encoded by clone lam |
| A_Geneseq_36:W45395 | + | 266.50 | 455.09 | 3.0e-18 | 237 | Mature prostate-specific glandu |
| A_Geneseq_36:W49087 | + | 266.50 | 455.09 | 3.0e-18 | 237 | Mature human Kallikrein 2 (hk2) |
| A_Geneseq_36:W83212 | + | 266.50 | 455.09 | 3.0e-18 | 237 | Variant human Kallikrein 2 (hk2) |
| A_Geneseq_36:W83202 | + | 266.50 | 455.09 | 3.0e-18 | 237 | Prostate-specific glandular kal |
| A_Geneseq_36:W96186 | + | 266.50 | 455.09 | 3.0e-18 | 237 | Prostate-specific glandular kal |
| A_Geneseq_36:W96190 | + | 266.50 | 455.05 | 3.0e-18 | 238 | Mature human Kallikrein 2 (hk2) |
| A_Geneseq_36:R84670 | + | 266.50 | 454.78 | 3.0e-18 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:R84669 | + | 266.50 | 454.78 | 3.0e-18 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W45396 | + | 266.50 | 454.78 | 3.0e-18 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W83204 | + | 266.50 | 454.78 | 3.0e-18 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:W96188 | + | 266.50 | 454.78 | 3.0e-18 | 244 | Prostate-specific glandular kal |
| A_Geneseq_36:R84668 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Prepro-hk2 kallikrein 2 (prohk2) |
| A_Geneseq_36:W06972 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Kallikrein prepro-hk2v21 varia |
| A_Geneseq_36:W06971 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W10600 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Human prepro-Trp226-glandular k |
| A_Geneseq_36:W45397 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W45400 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Prostate-specific glandular kal |
| A_Geneseq_36:W49085 | + | 266.50 | 454.07 | 3.1e-18 | 261 | Wild-type human Kallikrein 2 (h |

seq_name: A_Geneseq_36:W69389
seq_documentation_block:
ID W69389 standard; Protein; 164 AA.
AC W69389;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 118 /note= "unspecified amino acid"
FT
PN W09837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR N-PSDB; V58648.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 117-118; 14pp; English.
CC This sequence is encoded by a human prostatic tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 164 AA;

alignment_scores:
Quality: 854.00 Length: 164
Ratio: 5.239 Gaps: 0
Percent Similarity: 99.390 Percent Identity: 99.390

alignment_block:
US-09-030-606-177 x W69389 ..

Align seg 1/1 to: W69389 from: 1 to: 164

34 ATGGAAGAATGTTCTGCTGGGGCTCTGTCGATCCGAGTGGGT 83
|||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
84 GCTGTCAGCGGCACACTGTTCCAGAACTCTACACCATCGGGCTGGCC 133
|||||
17 LeuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
134 TGCACAGTCTTGAGGGCGACCAAGAGCCAGGAGCCAGATGTGGAGGCC 183
|||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
184 AGCCTTCGCTACGCGACCCAGAGTACAACAGACCCCTGCTGCTAACGA 233
|||||
51 SerLeuSerValArgHisProGluTyrAsnArgProLeuLeuAlaAsnAs 67
234 CCTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCATCC 283
|||||
67 pleuMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84
284 GGAGCATCAGCATTGCTTCGACAGTGCCTACCGCGGGGGAACCTTGCCCTC 333


```

|||||
84  rgSerileSerileAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
334 GTTCTGGCTGGGGCTGCTGCTGGCGAAGCATGCTGTGATTCGCATCCAGTC 383
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAlaIleGlnSe 117
384 CCAGACTGTGGAGGCTGGAGGTGTGAGAGCTTTCCCAACCCCTGGCAGG 433
|||||
117 r***ThrValGlyGlyTrpGlyLeuGlnCysGlnLysLeuSerGlnProTrpGlnG 134
434 GTTGTACCATTTTCGGCACTTCAGTGCAAGAGACGTCCTGTCATCCTC 483
|||||
134  lYcysThrIleSerAlaThrSerSerAlaArgThrSerCysCysIleLeu 150
484 ACTGGGTGCTCACTACTGCTCACTGCATCACCCGGACACTG 525
|||||
151 ThrGlyCysSerLeuLeuLeuThrAlaSerProGlyThrLeu 164

seq_name: A_Geneseq_36:W71873

```

```

seq_documentation_block:
ID  W71873 standard; Protein; 164 AA.
AC  W71873;
DT  06-JAN-1999 (first entry)
DE  Protein encoded by prostate tumour clone P703 splice variant DE14.
KW  Prostate; cancer; tumour; vaccine; immunogen; clone.
OS  Homo sapiens.
FH  Key
FT  MISC_difference 118
FT  Location/Qualifiers
FT  /note= "undefined residue"
PN  WO9837093-A2.
PD  27-AUG-1998.
PF  25-FEB-1998; U03492.
PR  09-FEB-1998; US-020956.
PR  25-FEB-1997; US-806099.
PR  01-AUG-1997; US-904804.
PA  (CORI-) CORIXA CORP.
PI  Dillon DC, Xu J;
DR  WPI; 98-609886/51.
PT  Polypeptides comprising immunogenic portions of prostate proteins -
PT  used in a vaccine for the treatment of prostate cancer
PS  Example 3; Page 109; 130pp; English.
CC  The present sequence is an immunogenic portion of a prostate tumour
CC  protein. The immunogen, or the DNA encoding it, can be used as a
CC  vaccine for the treatment of prostate cancer. The immunogen was
CC  isolated from a prostate tumour cDNA library obtained by subtracting
CC  a prostate tumour cDNA expression library with a normal tissue cDNA
CC  library.
SQ  Sequence 164 AA;

```

```

alignment_scores:
Quality: 854.00      Length: 164
Ratio: 5.239        Gaps: 0
Percent Similarity: 99.390      Percent Identity: 99.390

```

alignment_block:

US-09-030-606-177 x W71873 ..

Align seg 1/1 to: W71873 from: 1 to: 164

```

34 ATGGAAGATGTTCTGCTCGGGCGTCTGTCATCCGAGTGGGT 83
|||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
84 GCTGTCAAGCCGACACTGTGTTCCAGAACTCCTACACATCGGGCTGGGCC 133
|||||
17 IleuSerAlaAlaHisCysPheGlnAsnSerTrpThrIleGlyLeuGlyL 34
134 TGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGAGGCC 183
|||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50

```

```

184 AGCCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCTAACGA 233
|||||
51 SerLeuSerValArgHisProGluTrpAsnArgProLeuLeuAlaAsnAs 67
234 CCTCATGCTCATCAAGTTGGACGAATCCGTCCTCCAGTGTGACACCATCC 283
|||||
67 pLeuMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84
284 GGACATCAGCATTCGCTTCGCAGTGGCCCTACCGCGGGGAACCTTGCCCTC 333
|||||
84  rgSerileSerileAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
334 GTTCTGGCTGGGGCTGCTGCGAAGCATGCTGTGATTCGCATCCAGTC 383
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAlaIleGlnSe 117
384 CCAGACTGTGGAGGCTGGAGGTGTGAGAGCTTTCCCAACCCCTGGCAGG 433
|||||
117 r***ThrValGlyGlyTrpGlyLeuGlnCysGlnLysLeuSerGlnProTrpGlnG 134
434 GTTGTACCATTTTCGGCACTTCAGTGCAAGAGACGTCCTGTCATCCTC 483
|||||
134  lYcysThrIleSerAlaThrSerSerAlaArgThrSerCysCysIleLeu 150
484 ACTGGGTGCTCACTACTGCTCACTGCATCACCCGGACACTG 525
|||||
151 ThrGlyCysSerLeuLeuLeuThrAlaSerProGlyThrLeu 164

seq_name: A_Geneseq_36:W59129

```

seq_documentation_block:

ID W59129 standard; Protein; 232 AA.

AC W59129;

DT 11-SEP-1998 (first entry)

DE Homo sapiens Tub Interactor (hTI-1) protein.

KW serine protease; tub interactor; treatment; obesity; cachexia;

KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;

KW neurodegenerative disease; Alzheimer's disease; drug screening;

KW Parkinson's disease; Huntington's chorea; detection; diagnosis;

KW amyotrophic lateral sclerosis; spinocerebellar degeneration.

OS Homo sapiens.

FH Key

FT Region

FT 42

FT Location/Qualifiers

FT /note= "undefined amino acid"

PN WO9812302-A1.

PD 26-MAR-1998.

PF 05-SEP-1997; U15627.

PR 21-JUL-1997; US-897340.

PR 17-SEP-1996; US-715032.

PA (MILL-) MILLENNIUM PHARM INC.

PI Errada PR, Gimeno CJ;

DR WPI; 98-217246/19.

DR N-PSDB; V11855.

PT Tub Interactor genes

PT of obesity, cachexia, anorexia nervosa or related disorders e.g.

PT diabetes

PS Claim 28; Fig 1; 120pp; English.

CC The sequence is that encoding the Tub Interactor protein (hTI-1)

CC which is a putative serine protease. TI genes function

CC in biochemical pathways involved in weight control and

CC related disorders. The products can be used for treating

CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,

CC or a related disorder such as diabetes. The products can

CC also be used to modulate cell cycle progression and apoptosis.

CC They can be used for treating neurodegenerative diseases

CC which are characterised by apoptosis, including Alzheimer's

CC disease, Parkinson's disease, Huntington's chorea, amyotrophic

CC lateral sclerosis or spinocerebellar degenerations. The

CC products can also be used for detection, diagnosis and

CC drug screening.

CC Sequence 232 AA;

SQ

alignment_scores:

Quality: 621.00 Length: 154
Ratio: 4.669 Gaps: 3
Percent Similarity: 86.364 Percent Identity: 81.169

alignment_block:

US-09-030-606-177 x W59129 ..

Align seg 1/1 to: W59129 from: 1 to: 232

```
1 GCGCACTCGACGCCCTGGCAGCGCGCACTGCTCATGGAAGAATTGTT 50
|||||
5 AlaHisSerGlnProTrpGlnAlaAlaLeuValMetGluAsnGluLeuPh 21
51 CTGCTCGGGCGCTCTGTGTCATCCGACATGGGTGCTGTACGCCGCACT 100
|||||
21 ecysSerGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisC 38
101 GTTTC.....CAGACTCCTACACCATCGGGCTGGGCTGCAC 138
|||||
38 yspheGlnLys**ValGlnSerSerTyrThrIleGlyLeuGlyLeuHis 54
139 AGCTTGAGGGCCGACCAAGACCGCAGGAGCGAGATGGTGAGGCCACT 188
|||||
55 SerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAlaSerle 71
189 CTCCGTACGGCAGCCAGATACAACAGACCCCTGCTGCTAACGACCTCA 238
|||||
71 uSerValArgHisProGluTyrAsnArgProLeuLeuAlaAsnAspLeuM 88
239 TGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGAGC 288
|||||
88 etLeuIleLysLeuAspGluSerValSerGluSerAspThrIleArgSer 104
289 ATCAGCATGCTTCGACAGTGCCTTACCGCGGGGAACCTTGCTGCTGTT 338
|||||
105 IleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeuValSe 121
339 TGGCTGGGGTCTGCTGGCGAAGCATGCTGTG..... 369
|||||
121 rGlyTrpGlyLeuLeuAlaAsnGlyArgMetProThrValLeuGlnCysV 138
370 .....ATTGCCATCCAGTCCAGACTGTGGCAGAGGCTGGAGTGTGAGA 414
|||||
138 alAsnValSerValValSerGluGluVal.....CysSerLys 150
415 CTTCCCAACCC 426
|||||
151 LeuTyrAspPro 154
```

seq_name: A_Geneseq_36:W60592

seq_documentation_block:

ID W60592 standard; Protein; 248 AA.

AC W60592;

DT 07-SEP-1998 (first entry)

DE Human prostate-specific kallikrein (HPSK) protein.

KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;

KW benign prostate hyperplasia; diagnosis; drug screening; PSK.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 113

FT /label= unknown

FT /note= "encoded by NTC"

FT Misc_difference 128

FT /label= unknown

FT /note= "encoded by AGN"

FT Misc_difference 132

FT /label= unknown

FT /note= "encoded by GNT"

PN W09820117-A1.

PD 14-MAY-1998.

PF 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goll SK;
DR WPL; 98-286933/25.
DR N-PSDB; V37495.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 1; Fig 1A-C; 68pp; English.
CC This represents a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 248 AA;

alignment_scores:

Quality: 610.00 Length: 149
Ratio: 4.692 Gaps: 2
Percent Similarity: 87.248 Percent Identity: 80.537

alignment_block:

US-09-030-606-177 x W60592 ..

Align seg 1/1 to: W60592 from: 1 to: 248

```
4 CACTCGCAGCCCTGGCAGCGCGCACTGCTCATGGAAGAATGTTCTG 53
|||||
34 HisSerGlnProTrpGlnAlaAlaLeuValMetGluAsnGluLeuPheC 50
54 CTGCGGCGCTCTGTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 103
|||||
50 sSerGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysP 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 153
|||||
67 heGlnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAsp 83
154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACC 203
|||||
84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisPr 100
204 AGAGTACAACAGACCCTTGCTGCTGCTAACGACCTCATGCTCATAGTTG 253
|||||
100 oGluTyrAsnArgProLeuLeuAlaAsnAspLeuMet**IleLysLeuA 117
254 ACGAATCCGTGTCGAGTCTGACACCATCCGAGCATGCAATGCTTGC 303
|||||
117 spGluSerValSerGluSerAspAsnIleArg**IleSerIle**Ser 133
304 CAGTGCCCTACCGCGGGGGAACCTTGCTGCTGTTCTGCTGGGTCTGCT 353
|||||
134 GlnCysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLe 150
354 GCGGACGATGCTGTG.....ATTGCCATCC 379
|||||
150 uAlaAsnGlyArgMetProThrValLeuGlnCysValAsnValSerValV 167
380 AGTCCAGACTGTGGAGGCTGGAGTGTGAGAAGCTTCCCAACCC 426
|||||
167 alSerGluGluVal.....CysSerLysLeuTyrAspPro 178
```

seq_name: A_Geneseq_36:W69388

seq_documentation_block:

ID W69388 standard; Protein; 205 AA.

AC W69388;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key
FT Misc_difference 127
FT /note= "unspecified amino acid"
FT Misc_difference 204
FT /note= "unspecified amino acid"
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR N-PSDB; V58647.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 115-116; 14ipp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 205 AA;

alignment_scores:
Quality: 569.00 Length: 139
Ratio: 4.664 Gaps: 2
Percent Similarity: 87.770 Percent Identity: 82.014

alignment_block:
US-09-030-606-177 x W69388 ..

Align seg 1/1 to: W69388 from: 1 to: 205

34 ATGGAAGAAGATGTCTGCTCGGGCGTCTGTCATCCGAGTGGGT 83
|||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
84 GCTGTACGCCGACACTGTTTCCAGAACTCTACACATCGGGCTGGGCC 133
|||||
17 IleuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
134 TGCACAGTCTTGAGGGCCGACCAAGAGCCAGGGAGCCAGTGTGAGGCC 183
|||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
184 AGCCTCTCCGTACGGCAGCCAGAGTACAACAGACCCCTTGCTCGCTAACGA 233
|||||
51 SerLeuSerValArgHisProGluTyrAsnArgLeuLeuAlaAsnAs 67
234 CCTCATGCTCATCAAGTGGAGGAATCCGTGTCCGAGTCTGACACCATCC 283
|||||
67 pleuMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84
284 GGAGCATCAGCATGTGCTTCGAGTGCCTTACCGCGGGGGAACCTTGCCCTC 333
|||||
84 rGSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
334 GTTCTGTGGCTGGGCTGTGCTGGCGAAGATGCTGTG..... 369
|||||
104 ValSerGlyTyrGlyLeuLeuAlaAsnGlyArgMetProThrValLeuHi 117

370ATTGCCAATCCAGTCCAGACTGTGGAGGCTGGAGTGTG 409
:::|||||:::|||||
117 scysValasnValSerValValSerGlu**Val.....CysS 130
410 AGAAGCTTTCACCAACC 426
:::|||||:::|||||
130 erLysLeuTyrAspPro 135

seq_name: A_Geneseq_36:W71872

seq_documentation_block:

ID W71872 standard; Protein; 205 AA.
AC W71872;
DT 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key
FT Misc_difference 127
FT /note= "undefined residue"
FT Misc_difference 204
FT /note= "undefined residue"
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Example 3; Page 107-108; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 205 AA;

alignment_scores:
Quality: 569.00 Length: 139
Ratio: 4.664 Gaps: 2
Percent Similarity: 87.770 Percent Identity: 82.014

alignment_block:
US-09-030-606-177 x W71872 ..

Align seg 1/1 to: W71872 from: 1 to: 205

34 ATGGAAGAAGATGTCTGCTCGGGCGTCTGTCATCCGAGTGGGT 83
|||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
84 GCTGTACGCCGACACTGTTTCCAGAACTCTACACATCGGGCTGGGCC 133
|||||
17 IleuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
134 TGCACAGTCTTGAGGGCCGACCAAGAGCCAGGGAGCCAGATGTGAGGCC 183
|||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
184 AGCCTCTCCGTACGGCAGCCAGAGTACAACAGACCCCTTGCTCGCTAACGA 233
|||||
51 SerLeuSerValArgHisProGluTyrAsnArgLeuLeuAlaAsnAs 67
234 CCTCATGCTCATCAAGTGGAGGAATCCGTGTCCGAGTCTGACACCATCC 283
|||||
67 pleuMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84


```
284 GGAGCATCAGCATTCGTCGACATGCCCCCTACCCGGGGAACTCTTGCCTC 333
      |||||||
      84 rgsrileSerileAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
334 GTTCTGGCTGGGGCTGCTGCGGACGATGCTGTG..... 369
      |||||||
      101 ValSerGlyTrpGlyLeuAlaAsnGlyArgMetProThrValLeuH1 117
370 .....ATTGCCATCCAGTCCCGACACTGTGGAGGCTGGAGTGTG 409
      :::::
      117 scysValAsnValSerValSerValSerGlu**Val.....CysS 130
410 AGAGCTTTCCCAACC 426
      ::|||
      130 erLysLeuTyrAspPro 135

seq_name: A_Geneseq_36:W69387
```

```
seq_documentation_block:
ID W69387 standard; Protein; 159 AA.
AC W69387;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DEI protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
   therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 103 /note= "unspecified amino acid"
FT Misc_difference 105 /note= "unspecified amino acid"
FT W09837418-A2.
PN 27-AUG-1998.
PD 25-FEB-1998; U03690.
PF 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR N-PSDB; V58644.
PT Novel human prostate specific tumour protein and fragments - useful
   for detecting and treating prostate cancers
PS Example 1; Page 112-113; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
   can be used in the method of the invention. The method is for detecting
   prostate cancer comprises contacting a biological sample with an agent
   able to bind an immunogenic portion of a prostate protein (such as
   this protein sequence). An antibody which binds to an immunogenic
   portion of the prostate protein, and the method can be used to detect,
   monitor progression of, or treat prostate cancers. The antibody may
   also be conjugated to a therapeutic agent for use in therapy of prostate
   cancers.
SQ Sequence 159 AA;
```

```
alignment_scores:
Quality: 328.00 Length: 93
Ratio: 4.316 Gaps: 2
Percent Similarity: 81.720 Percent Identity: 74.194
```

alignment_block:

US-09-030-606-177 x W69387 ..

Align seg 1/1 to: W69387 from: 1 to: 159

```
172 ATGTGGAGGCGCAGCCTCTCCGTACGGCACCCAGAGTACAAGACCCTT 221
      |||||||
      1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17
222 GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGT 271
      |||||||
```

```
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
272 CTGACACCATCCGAGACATTCAGCATTCGTCGAGTGCCCTACCGGGG 321
      |||||||
      34 erspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
322 AACTCTGCCCTGTTCTGGCTGGGGTCTGCTGGCCGACGATGCTGTG.. 369
      |||||||
      51 AsnSerCysLeuValSerGlyTrpGlyLeuAlaAsnGlyArgMetPr 67
370 .....ATTGCCATCCAGTCCCGACACTGTGGAG 397
      :::::
      67 oThrValLeuGlnCysValAsnValSerValSerGluGluVal.... 82
398 GCTGGAGTGTGAGAGCTTTCCCAACC 426
      |||:::
      83 .....CysSerLysLeuTyrAspPro 89

seq_name: A_Geneseq_36:W71871
```

```
seq_documentation_block:
ID W71871 standard; Protein; 159 AA.
AC W71871;
DT 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DEI.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 103 /note= "undefined residue"
FT Misc_difference 105 /note= "undefined residue"
FT W09837093-A2.
PN 27-AUG-1998.
PD 25-FEB-1998; U03492.
PF 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
   used in a vaccine for the treatment of prostate cancer
PS Example 3; Page 105; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
   protein. The immunogen, or the DNA encoding it, can be used as a
   vaccine for the treatment of prostate cancer. The immunogen was
   isolated from a prostate tumour cDNA library obtained by subtracting
   a prostate tumour cDNA expression library with a normal tissue cDNA
   library.
SQ Sequence 159 AA;
```

```
alignment_scores:
Quality: 328.00 Length: 93
Ratio: 4.316 Gaps: 2
Percent Similarity: 81.720 Percent Identity: 74.194
```

alignment_block:

US-09-030-606-177 x W71871 ..

Align seg 1/1 to: W71871 from: 1 to: 159

```
172 ATGTGGAGGCGCAGCCTCTCCGTACGGCACCCAGAGTACAAGACCCTT 221
      |||||||
      1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17
222 GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGT 271
      |||||||
      17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
272 CTGACACCATCCGAGACATTCAGCATTCGCTGCGAGTGCCCTACCGGGG 321
      |||||||
```


34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
322 AACTCTTGCCCTGCTTCTTGCGTGGGCTGCTGGCGAAGATGCTGTG.. 369
|||||
51 AsnSerCysLeuValSerGlyTrpGlyLeuLeuAlaSnGlyArgMetPr 67
370ATTGCCATCCAGTCCCGAGACTGTGGAG 397
:::|::|::|
67 OThrValLeuGlnCysValAsnValSerValValSerGluGluVal.... 82
398 GCTGGAGTGTGAGAGCTTCCCAACCC 426
|||::|::|::|
83CysSerLysLeuTyrAspPro 89

seq_name: A_Geneseq_36:Y12281

seq_documentation_block:

ID Y12281 standard; Protein; 66 AA.
AC Y12281;
DT 17-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO:312.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Ducleit A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR N-PSDB; X41114.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 27; Page 661; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 66 AA;

alignment_scores:

Quality: 318.00 Length: 63
Ratio: 5.048 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-030-606-177 x Y12281 ..

Align seg 1/1 to: Y12281 from: 1 to: 66

172 ATGTTGAGGCCAGCTCTCCGTACGGCAGCCAGAGTACACAGACCCCTT 221
|||||
..1MetValGlnValSerLeuSerValArgHisProGluTyrAsnArgProLe 17

222 GCTGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGT 271
|||||
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGlnSerValSerGln 34
272 CTGACACCATCCGGAGCATCAGCATTTGCTTCGACAGTGCCTTACCGCGGG 321
|||||
34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
322 AACTCTTGCCCTGCTTCTTGCGTGGGCTGCTGGCGAAC 360
|||||
51 AsnSerCysLeuValSerGlyTrpGlyLeuLeuAlaSn 63

seq_name: A_Geneseq_36:W10694

seq_documentation_block:

ID W10694 standard; Protein; 260 AA.
AC W10694;
DT 08-APR-1997 (first entry)
DE Human recombinant neuropsin, used for antibody production.
KW Antibody; Alzheimer's disease; neurodegenerative; epitope; epilepsy;
KW recombinant.
OS Homo sapiens.
FH Key
FT peptide
FT location/Qualifiers
FT 95..215
FT /note="claimed fragment of the protein, used for
FT antibody production"
PN J08245700-A.
PD 24-SEP-1996.
PF 14-MAR-1995; 083154.
PR 14-MAR-1995; JP-083154.
PA (IGAK-). IGAKU SEIBUTSUGAKU KENKYUSHO KK.
DR WPI: 96-482259/48.
DR N-PSDB; T48519.
PT Antibody against neuropsin - used in the treatment of Alzheimer's
PT disease and epilepsy
PS Claim 1; Page 7; 9pp; Japanese.
CC W10694 is a recombinantly produced human neuropsin protein. The
CC main invention relates to an antibody against the neuropsin protein,
CC in particular a claimed fragment of the neuropsin protein (see W10695
CC and features table). The antibody can be used to treat
CC neurodegenerative disorders such as Alzheimer's disease and epilepsy.
SQ Sequence 260 AA;

alignment_scores:

Quality: 296.00 Length: 172
Ratio: 2.446 Gaps: 5
Percent Similarity: 70.349 Percent Identity: 35.465

alignment_block:

US-09-030-606-177 x W10694 ..

Align seg 1/1 to: W10694 from: 1 to: 260

4 CACTGCGACCCCTGGCAGCGGCTGCTCATGGAACGAATGTTCTG 53
|||||
42 HisserGlnProTrpGlnAlaAlaLeuPheGlnGlyGluArgLeuIleCy 58
54 CTGGGCGCTCTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 103
|:::|
58 sGlyGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysL 75
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGCCGAC 153
:::|::|
75 ySLysGlnLysTyrSerValArgLeuGlyAspHisSerLeuGlnSerArg 91
154 CAAGAGCCAGGAGCAGCATGATGTGAGAGCCAGCCTCTCCGTACGGCACC 203
:::|::|
92 AspGlnPro...GluGlnGluIleGlnValAlaGlnInsertIleGlnHisPr 107
204 AGAGTACAACAGACCCCTTG.....CTGCTAACGACCTCATGCTCA 244
|
|||||

107 OCSTYrAsnAsnSerAsnProGluAspHisSerHisAspIleMetLeuI 124
245 TCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGC 294
124 leArgLeuGlnAsnSerAlaAsnLeuGlyAspIysValIysProValGln 140
295 ATTGCTTCGAGTGCCTTACCGCGGGGAACTCTGCTCGTTCTTGCTG 344
141 leuAlaAsnLeuCysProIysValGlyGlnIysCysIleIleSerGlyTr 157
345 GGGTCTGCTGGCG.....AACGATG 364
157 pGlyThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysA 174
365 CTGTGATTGCCATCCAGTCCAGACTGTGGAGGCTGGAGTGTAGAAG 414
174 laGluValIysIleTyrSerGlnAsn.....LysCysGluArg 186
415 CTT.....TCCCAACCCCTGGCAGGGTTGTACCATTTCCGCAACTTCAG 458
187 AlaTyrProGlyIysIleThrGluGlyMetValCysAlaGlySerSerAs 203
459 TGC AAGGAGCTCCTGC 474
203 nGlyAlaAspThrCys 208

seq_name: A_Geneseq_36:W12393

seq_documentation_block:

ID W12393 standard; Protein; 260 AA.
AC W12393;
DT 15-MAY-1997 (first entry)
DE Mouse neuropilin protein.
KW Mouse; neuropilin; hippocampus; lambda gt10; primer; PCR; amplification;
KW polymerase chain reaction; serine protease domain; nerve growth factor;
KW NGF; insect cell; virus; expression vector; transfection;
KW cerebral disease.
OS Mus musculus.
PN J08311099-A.
PD 26-NOV-1996.
PF 13-MAR-1996; 056367.
PR 14-MAR-1995; JP-054584.
PA (SHIO/) SHIOZAKA S.
DR WPI; 97-061812/06.
DR N-PSDB; T63251.
PT Nucleic acid encoding neuropilin - for producing neuropilin, useful
PT for diagnosis and treatment of cerebral disease
PS Claim 1; Page 6-7; 9pp; Japanese.
CC This is the amino acid sequence of a novel mouse protein designated
CC neuropilin. The encoding gene was isolated from a mouse hippocampal
CC cDNA library in lambda gt10 using a cloned, amplified fragment of
CC the gene (clone B41; T63254). This fragment was amplified using
CC primers T63252-3. The primers were synthesised based on the serine
CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated
CC 6 positive clones, of which clone NP5 contained the longest insert
CC (this sequence). The protein has a molecular weight of around 26 kD.
CC It has 43% homology with EGF-BP, 41 % with NGF-gamma; 39% with NGF-alpha;
CC 38% with trypsin and 18% with tPA. The protein can be used for clinical
CC diagnosis and treatment of cerebral diseases.
SQ Sequence 260 AA;

alignment_scores:

Quality: 296.00 Length: 172
Ratio: 2.446 Gaps: 5
Percent Similarity: 70.349 Percent Identity: 35.465

alignment_block:

US-09-030-606-177 x W12393 ..

Align seg 1/1 to: W12393 from: 1 to: 260

4 CACTCGCAGCCCTGGCAGCGGCGCACTGTCATGAAACGAATGTCTCTG 53

||||| 154 CAAGAGCCAGGAGCAGATGCTGAGGCCAGCCTCTCCGTACGGCACC 203
42 HisSerGlnProTrpGlnAlaAlaLeuPheGlnGlyGluArgLeuIleCy 58
54 CTGGGGCGTCTGTGTCATCCGAGTGGGTGTGTCACCGGCACACTGTT 103
58 sGlyGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysL 75
104 TCCAGAACTCTACACCATCGGCGGTGGCCTGCACAGTCTTGAGCCGAC 153
75 yslYsGlnIysTyrSerValArgLeuGlyAspHisSerLeuGlnSerArg 91
154 CAAGAGCCAGGAGCAGATGCTGAGGCCAGCCTCTCCGTACGGCACC 203
92 AspGlnPro...GluGlnGluIleGlnValAlaGlnSerIleGlnHisPr 107
204 AGAGTACACAGACCCTTG.....CTGCTAACGACCTCATGCTCA 244
107 OCSTYrAsnAsnSerAsnProGluAspHisSerHisAspIleMetLeuI 124
245 TCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGC 294
124 leArgLeuGlnAsnSerAlaAsnLeuGlyAspIysValIysProValGln 140
295 ATTGCTTCGAGTGCCTTACCGCGGGGAACTCTGCTCGTTCTTGCTG 344
141 leuAlaAsnLeuCysProIysValGlyGlnIysCysIleIleSerGlyTr 157
345 GGGTCTGCTGGCG.....AACGATG 364
157 pGlyThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysA 174
365 CTGTGATTGCCATCCAGTCCAGACTGTGGAGGCTGGAGTGTAGAAG 414
174 laGluValIysIleTyrSerGlnAsn.....LysCysGluArg 186
415 CTT.....TCCCAACCCCTGGCAGGGTTGTACCATTTCCGCAACTTCAG 458
187 AlaTyrProGlyIysIleThrGluGlyMetValCysAlaGlySerSerAs 203
459 TGC AAGGAGCTCCTGC 474
203 nGlyAlaAspThrCys 208

seq_name: A_Geneseq_36:W94493

seq_documentation_block:

ID W94493 standard; Protein; 268 AA.
AC W94493;
DT 23-APR-1999 (first entry)
DE Human kallikrein.
KW Human; kallikrein; keratinocyte; HKAL; skin disorder; cancer; eczema;
KW psoriasis; scleroderma; adenocarcinoma; leukaemia; melanoma.
OS Homo sapiens.
PN W09842849-A1.
PD 01-OCT-1998.
PF 25-MAR-1998; U05939.
PR 26-MAR-1997; US-824874.
PA (INCY-) INCYTE PHARM INC.
PI Hillman JL, Lal P;
DR WPI; 99-070073/06.
DR N-PSDB; X16295.
PT Human kallikrein polypeptide. HKAL - useful e.g. to treat skin
PT disorders e.g. eczema, psoriasis and screen for antagonists useful
PT to treat skin disorders and cancers
PS Claim 1; Page 41-42; 61pp; English.
CC The present sequence represents human kallikrein, designated HKAL.
CC HKAL (or fragments) can be administered therapeutically to increase
CC proteolysis and subsequent skin scaling to treat/prevent skin disorders,
CC e.g. eczema, psoriasis and scleroderma. HKAL has chemical and
CC structural homology with human stratum corneum chymotryptic enzyme
CC (thought to be involved in the proteolysis of intercellular cohesive
CC structures necessary for desquamation, the process by which outer layers
CC of skin are eliminated), and its expression in cDNA libraries was

CC associated with tumour-associated tissues and skin cells. It can be used
CC to screen for antagonists and agonists, and to generate antibodies.
CC HKaLl antagonists can be used to suppress excessive proteolysis and
CC subsequent skin cell scaling, so may be administered to treat skin
CC disorders. They may also be used to suppress excessive cell
CC proliferation, so can be administered to treat/prevent cancer, e.g.
CC adenocarcinoma, leukaemia and melanoma. Antibodies specific for HKaLl
CC may be used directly as antagonists, or indirectly as a targeting or
CC delivery mechanism for bringing pharmaceutical agents to HKaLl-expressing
CC cells. They are also useful to diagnose conditions/diseases characterised
CC by HKaLl expression and to monitor therapeutic interventions. The
CC polynucleotide encoding HKaLl, or complementary sequences, can be used to
CC produce hybridisation probes, useful to detect polynucleotides encoding
CC HKaLl, e.g. to diagnose diseases relating to polypeptide expression
CC (e.g. cancers of the bladder, prostate) or monitor HKaLl regulation
CC during therapeutic intervention. Polynucleotides encoding HKaLl are
CC useful to produce antisense sequences for therapeutic administration to
CC modulate/prevent HKaLl expression e.g. to treat/prevent skin disorders
CC or cancer as above.
SQ Sequence 268 AA;

alignment_scores: Quality: 292.50 Length: 142
 Ratio: 2.985 Gaps: 3
Percent Similarity: 69.014 Percent Identity: 42.254

alignment_block:
US-09-030-606-177 x W94493 ..

Align seg 1/1 to: W94493 from: 1 to: 268

```
19 CAGCGGGCACTGGTCATGAA...AAGCAATGTCTGCTCGGGCTCTCT 65
   |||||
56 GlnAlaAlaLeuLeuLeuArgProAsnGlnLeuTyrcysGlyAlaValLe 72
66 GGTGCATCCGAGTGGTGTCTGTCAGCCGACACACTGTTCCAGAACTCT 115
   |||||
72 uValHisProGlnTrpLeuLeuThrAlaAlaHisCysArgLysLysValP 89
116 ACACCATCGGGCTGGGCTGCAGACAGTCTTGAGCGCCAGCAAGCCAGG 165
   ::::
89 heargValArgLeuGlnHisTyrrSerLeuSerProValTyrrLuserGly 105
166 AGCAGATGCTGGAGGCCAGCCTCTCCGTACGGCAGCCAGATACAGAG 215
   ::|||
106 GlnGlnMetPheGlnGlnValLysSerIleProHisProGlyTyrrSerH 122
216 ACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTGGACGAATCCGT 265
   ::|||
122 sProGlnHisSerAsnAspLeuMetLeuIleLysLeuAsnArgArgIleA 139
266 CCGAGTCTGACACCATCCGAGCATCAGCATGCTTCGAGTGCCTTACC 315
   ::::
139 rgrProThrLysAspValArgProIleAsnValSerSerHisCysProSer 155
316 GCGGGCACTCTTGCCTGTTCTGCTGGGCTGCTGGCGAAGCATGC 365
   |||||
156 AlaGlyThrLysCysLeuValSerGlyTrpGlyThrLysSerProG 172
366 TGTG.....ATGCCATCCAGTCCC 385
   |||
172 nValHisPheProLysValLeuGlnCysLeuAsnIleSerValLeuSerG 189
386 AGACTGTGGAGGCTGGAGTGTGAG 411
   ||
189 InLys.....ArgCysGln 193
```

seq_name: A_Geneseq_36:R67888

seq_documentation_block:
ID R67888 standard; Protein; 253 AA.
AC R67888

DT 09-AUG-1995 (first entry)
DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
OS Homo sapiens.
PN WO9500651-A.
PD 05-JAN-1995.
PF 20-JUN-1994; IB0166.
PR 18-JUN-1993; DK-000725.
PA (SYMB-) SYMBICOM AB.
PI Egelrud T, Hansson L;
DR WPI: 95-052088/07.
DR N-PSDB; Q81203.
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides,
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
PS Disclosure; Page 97; 137pp; English.
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
SQ Sequence 253 AA;

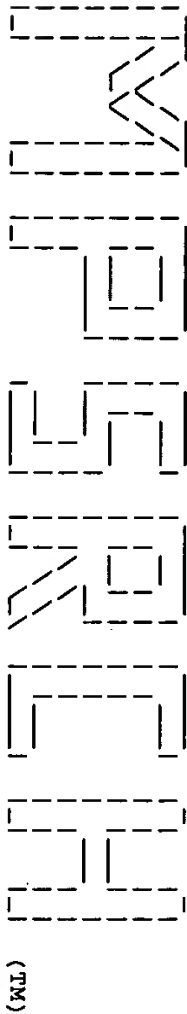
alignment_scores: Quality: 274.50 Length: 143
 Ratio: 2.745 Gaps: 3
Percent Similarity: 69.930 Percent Identity: 38.462

alignment_block:
US-09-030-606-177 x R67888 ..

Align seg 1/1 to: R67888 from: 1 to: 253

```
7 TCGCAGCCCTGGCAGCGCCACTGTGTCATGGAAGAAGCAATGTTCTGCTC 56
   |||||
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG 56
57 GGGCTCTGTGTCATCCGAGTGGGCTGCTGTCAGCCGACACTGTTCC 106
   |||||
56 GlyValLeuValAsnGlnArgTrpValLeuThrAlaAlaHisCysLysM 73
107 AGAACTCTACACCATCGGGCTGGGCTGCAGACAGTCTTGAGCGCCAGCA 156
   ::|||
73 eValAsnGlnTrpThrValHisLeuGlySerAspThrLeu...GlyAspArg 88
157 GAGCCAGGAGCCAGATGCTGGAGGCCAGCCTCTCCGTACGGCAGCCAGA 206
   ::::
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG 103
207 GTACAACAGACCTTGCCTGCTAACGACCTCATGCTCATCAAGTTGGACG 256
   |||||
103 yTyrrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsn 120
257 AATCGGTGTCCGAGTGTGACACCATCCGAGCATCAGCATGCTTCGACAG 306
   ::|||
120 eGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
307 TGGCCTTACCGGGGCACTTGTGCTGCTGTTCTGCTGGGCTGCTGCGC 356
   |||
137 CysGlnProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153
357 GAACGATGCTGTG.....ATGCCATCCAGTCCAGCA 388
   ::|||
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170
389 CTGTGGAGGCTGGAGTGTGAGAGCTT 417
   ::|||
170 euLeuSerProGlnAspCysThrLysVal 179
```


104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
:::||||| :: ||||| ||::|||
67 leSerAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu..Pheasp 82
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCAGCC 203
:::||||| ::|||:::||||| |||
83 AspGluAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr 99
204 AGAGTACACAGACCCCTGCTCGCTAAC..... 231
| ::||| ||||| |||
99 oglyPheAsnMetSerLeuLeuGluAsnHisThrArgGlnAlaAspGluA 116
232GACCTCATGCTCATCAAGTTGAGCAATCCGTGCCGAG 270
||||| |||||:::||||| ||| :::::
116 sPtyrSerHisAspLeuMetLeuLeuArgLeuThrGluProAlaAspThr 132
271 ...TCTGACACCATCCGAGCATCAGCATTCGCTGCAGTGCCCTACCGC 317
:::|||||::: ::::: |||
133 IleThrAspAlaValValValGluLeuProThrGlnGluProGluVal 149
318 GGGGAACCTCTGCCCTCGTTCTGGCTGGGCTCTGCTGCCG..... 357
:||||:::|||||:::||||| |||
149 IglySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPr 166
358AACGATGCTGTGATTGCCATCCAGTCCAGACGACTGTGGGAGGC 399
||| :: ::::: :::
166 heserPheProAspAspLeuGlnCysValAspLeuLysIleLeuProAsn 182
400 TGGGAGTGTGAGAG 414
|||||
183 AspGluCysGluLys 187



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 29 14:08:30 1999; Maspar time 11.40 Seconds

Tabular output not generated. 177.112 Million cell updates/sec

Title: >US-08-931-858E-174
Description: (1-37) from US08931858E.pep
Perfect Score: 265
Sequence: 1 LDLOEASVADKLSFGKMAETRGITWPHQGNHVR LPR 37

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: splrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 32.468; Variance 50.260; scale 0.646

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-----------|------------------------|-----------|
| 1 | 265 | 100.0 | 156 | 11 070300 | PERSEPHIN. | 1.66e-43 |
| 2 | 194 | 73.2 | 156 | 11 070301 | PERSEPHIN. | 2.85e-26 |
| 3 | 81 | 30.6 | 119 | 7 031201 | MHC H-2 S-REGION C4-SL | 5.10e-02 |
| 4 | 81 | 30.6 | 594 | 11 062238 | SEX-LIMITED PROTEIN SL | 5.10e-02 |
| 5 | 78 | 29.4 | 1464 | 10 082061 | PROTEIN INVOLVED IN ST | 1.70e-01 |
| 6 | 77 | 29.1 | 344 | 5 076253 | TRANSPOSASE. | 2.53e-01 |
| 7 | 77 | 29.1 | 345 | 5 061447 | TRANSPOSASE. | 2.53e-01 |
| 8 | 77 | 29.1 | 345 | 5 061451 | TRANSPOSASE. | 2.53e-01 |
| 9 | 77 | 29.1 | 345 | 5 061453 | TRANSPOSASE. | 2.53e-01 |
| 10 | 77 | 29.1 | 345 | 5 061452 | TRANSPOSASE. | 2.53e-01 |
| 11 | 77 | 29.1 | 345 | 5 061455 | TRANSPOSASE. | 2.53e-01 |
| 12 | 77 | 29.1 | 345 | 5 062616 | TRANSPOSASE. | 2.53e-01 |
| 13 | 77 | 29.1 | 345 | 5 061454 | TRANSPOSASE. | 2.53e-01 |
| 14 | 77 | 29.1 | 345 | 5 061446 | TRANSPOSASE. | 2.53e-01 |
| 15 | 77 | 29.1 | 345 | 5 061450 | TRANSPOSASE. | 2.53e-01 |
| 16 | 77 | 29.1 | 345 | 5 076252 | TRANSPOSASE. | 2.53e-01 |
| 17 | 77 | 29.1 | 553 | 10 004559 | T7N9.12. | 2.53e-01 |
| 18 | 77 | 29.1 | 1148 | 11 061372 | C4-COMPLEMENT PROTEIN | 2.53e-01 |
| 19 | 77 | 29.1 | 1738 | 11 070346 | COMPLEMENT C4. | 2.53e-01 |
| 20 | 74 | 27.9 | 266 | 10 080337 | ETHYLENE RESPONSIVE EL | 8.14e-01 |

| | | | | | | |
|----|----|------|------|-----------|------------------------|----------|
| 21 | 73 | 27.5 | 330 | 11 062353 | COMPLEMENT COMPONENT (| 1.19e+00 |
| 22 | 72 | 27.2 | 379 | 1 027636 | CDC6 RELATED PROTEIN. | 1.75e+00 |
| 23 | 72 | 27.2 | 577 | 10 049799 | SUCROSE-PHOSPHATE SYNT | 1.75e+00 |
| 24 | 71 | 26.8 | 462 | 2 005433 | ATP SYNTHASE SUBUNIT B | 2.54e+00 |
| 25 | 71 | 26.8 | 479 | 2 056670 | MANNOSE-SENSITIVE HEMA | 2.54e+00 |
| 26 | 71 | 26.8 | 880 | 3 000170 | REGULATORY PROTEIN. | 2.54e+00 |
| 27 | 70 | 26.4 | 107 | 14 067832 | DNA POLYMERASE (EC 2.7 | 3.69e+00 |
| 28 | 70 | 26.4 | 434 | 5 022170 | SIMILAR TO HUMAN ER-CO | 3.69e+00 |
| 29 | 70 | 26.4 | 845 | 14 067895 | DNA POLYMERASE (EC 2.7 | 3.69e+00 |
| 30 | 70 | 26.4 | 845 | 14 091533 | POL. PROTEIN. | 3.69e+00 |
| 31 | 69 | 26.0 | 139 | 2 054621 | DSPF. | 5.34e+00 |
| 32 | 68 | 25.7 | 265 | 10 040174 | MYB-RELATED PROTEIN. | 7.69e+00 |
| 33 | 68 | 25.7 | 367 | 10 004192 | MYB ISOLOG. | 7.69e+00 |
| 34 | 68 | 25.7 | 395 | 14 090337 | PHOSPHOPROTEIN. | 7.69e+00 |
| 35 | 68 | 25.7 | 488 | 5 044317 | REVERSE TRANSCRIPTASE | 7.69e+00 |
| 36 | 68 | 25.7 | 570 | 3 012457 | ORF YDR026C. | 7.69e+00 |
| 37 | 68 | 25.7 | 638 | 4 060317 | KIAA0571 PROTEIN. | 7.69e+00 |
| 38 | 68 | 25.7 | 2829 | 13 P70039 | ADENOMATOUS POLYPOSIS | 7.69e+00 |
| 39 | 67 | 25.3 | 306 | 2 007152 | PUTATIVE OXIDOREDUCTAS | 1.10e+01 |
| 40 | 67 | 25.3 | 461 | 5 017327 | AVERMECTIN-SENSITIVE G | 1.10e+01 |
| 41 | 67 | 25.3 | 495 | 5 017793 | FLI15.10 PROTEIN. | 1.10e+01 |
| 42 | 67 | 25.3 | 618 | 10 064378 | PUTATIVE TOPOISOMERASE | 1.10e+01 |
| 43 | 67 | 25.3 | 1007 | 4 043379 | R34001.1. | 1.10e+01 |
| 44 | 67 | 25.3 | 1676 | 10 023332 | CENTROMERE PROTEIN HOM | 1.10e+01 |
| 45 | 66 | 24.9 | 698 | 14 067889 | DNA POLYMERASE (EC 2.7 | 1.58e+01 |

ALIGNMENTS

| RESULT | ID | 1 | PRELIMINARY; | PRT; | 156 AA. |
|---|---|--|--------------|---------|---------|
| AC | 070300; | | | | |
| DT | 01-AUG-1998 (TREMBLREL. 07, CREATED) | | | | |
| DT | 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE) | | | | |
| DT | 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE) | | | | |
| DE | PERSEPHIN. | | | | |
| GN | PSPN. | | | | |
| OS | MUS MUSCULUS (MOUSE). | | | | |
| OC | EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA; | | | | |
| OC | SCIUROGNATHI; MORIDAE; MORINAE; MUS. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-129/SVT; | | | | |
| RX | MEDLINE; 98150950. | | | | |
| RA | MILBRANDT J., DE SAUVAGE F.J., FAHRNER T.J., BALOH R.H., LEITNER M.L., | | | | |
| RA | TANSEY M.G., LAMPE P.A., HEUCKEROH R.O., KOTZBAUER P.T., | | | | |
| RA | SIMBURGER K.S., GOLDEN J.P., DAVIES J.A., VEJSADA R., KATO A.C., | | | | |
| RA | HYNES M., SHERMAN D., NISHIMURA M., WANG L.-C., VANDLEN R., MOFFAT B., | | | | |
| RA | KLEIN R.D., POULSEN K., GRAY C., GARCES A., HENDERSON C.E., | | | | |
| RA | PHILLIPS H.S., JOHNSON E.M.; | | | | |
| RT | "Persephin, a novel neurotrophic factor related to GDNF and | | | | |
| RT | neurturin."; | | | | |
| RL | NEURON 20:245-253(1998). | | | | |
| DR | EMBL; AF040960; G2935706; -. | | | | |
| DR | MGI; MGI:1201684; PSPN. | | | | |
| SQ | SEQUENCE 156 AA; 17030 MW; D5445157 CRC32; | | | | |
| Query Match | | | | | |
| Best Local Similarity 100.0%; Score 265; DB 11; Length 156; | | | | | |
| Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | | | |
| DB | 24 | LDLOEASVADKLSFGKMAETRGITWPHQGNHVR LPR 60 | | | |
| OY | 1 | LDLOEASVADKLSFGKMAETRGITWPHQGNHVR LPR 37 | | | |
| RESULT 2 | | | | | |
| ID | 070301 | PRELIMINARY; | PRT; | 156 AA. | |
| AC | 070301; | | | | |
| DT | 01-AUG-1998 (TREMBLREL. 07, CREATED) | | | | |
| DT | 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE) | | | | |
| DT | 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE) | | | | |
| DE | PERSEPHIN. | | | | |

OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98150950.
RA MILBRANDT J., DE SAUVAGE F.J., FAHRNER T.J., BALOH R.H., LEITNER M.L.,
RA TANSEY M.G., LAMPE P.A., HEUCKEROOTH R.O., KOTZBAUER P.T.,
RA SIMBURGER K.S., GOLDEN J.P., DAVIES J.A., VEJSADA R., KATO A.C.,
RA HYNES M., SHERMAN D., NISHIMURA M., WANG L.-C., VANDLEN R., MOFFAT B.,
RA KLEIN R.D., POULSEN K., GRAY C., GARCES A., HENDERSON C.E.,
RA PHILLIPS H.S., JOHNSON E.M.;
RT "Persephin, a novel neurotrophic factor related to GDNF and
RT neurturin.";
RL NEURON 20:245-253(1998).
DR EMBL; AF040961; G2935708; -
SQ SEQUENCE 156 AA; 17063 MW; BA9BA08B CRC32;

Query Match 73.2%; Score 194; DB 11; Length 156;
Best Local Similarity 78.4%; Pred. No. 2.85e-26;
Matches 29; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 24 LDLOEAPADELSSGKMAETGRTWKPHOGNNVRLPR 60
|||||: |||||: |||||: |||||: |||||:
QY 1 LDLOEASVADKLSFGKMAETRGTTWTPHOGNNHVRLLR 37

RESULT 3
ID Q31201 PRELIMINARY; PRT; 119 AA.
AC Q31201;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MHC H-2 S-REGION C4-SLP (SEX-LIMITED PROTEIN) (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85038859.
RA TOSI M., LEVI-STRAUSS M., DUPONCHEL C., MEO T.;
RT "Sequence heterogeneity of murine complementary DNA clones related to
RT the C4 and C4-Slp isoforms of the fourth complement component.";
RL PHLOS. TRANS. R. SOC. LOND., B, BIOL. SCI. 306:389-394(1984).
DR EMBL; K02799; G199626; -
DR PFAM; PF00207; A2M; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 12972 MW; 7302A10E CRC32;

Query Match 30.6%; Score 81; DB 7; Length 119;
Best Local Similarity 55.6%; Pred. No. 5.10e-02;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 18 GKMADKAASWLTHOGNFH 35
|||||: |||||: |||||: |||||:
QY 15 GKMAETRGTTWTPHOGNNH 32

RESULT 4
ID Q62238 PRELIMINARY; PRT; 594 AA.
AC Q62238;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SEX-LIMITED PROTEIN SLP(W7) ALPHA-GAMMA CHAIN (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B10.D2(C4(H)SLP(A));

RX MEDLINE; 86176748.
RA HEMENWAY C., KALFF M., STAVENHAGEN J., WALTHALL D., ROBINS D.;
RT "Sequence comparison of alleles of the fourth component of complement
RT (C4) and sex-limited protein (SLP).";
RL NUCLEIC ACIDS RES. 14:2539-2554(1986).
DR EMBL; X06454; G54106; -
DR PFAM; PF00207; A2M; 1.
FT NON_TER 1 1
FT CHAIN 1 303
FT CHAIN 304 594
SQ SEQUENCE 594 AA; 66454 MW; C76E7A07 CRC32;

Query Match 30.6%; Score 81; DB 11; Length 594;
Best Local Similarity 55.6%; Pred. No. 5.10e-02;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 131 GKMADKAASWLTHOGNFH 148
|||||: |||||: |||||: |||||:
QY 15 GKMAETRGTTWTPHOGNNH 32

RESULT 5
ID O82061 PRELIMINARY; PRT; 1464 AA.
AC O82061;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PROTEIN INVOLVED IN STARCH METABOLISM PRECURSOR.
GN R1.
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. DESIRE;
RA LOBERETH R., KOSSMAN J.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y09533; E303790; -
KW SIGNAL.
FT SIGNAL 1 48
FT SIGNAL 1464 AA; 163236 MW; 564CF92A CRC32;

Query Match 29.4%; Score 78; DB 10; Length 1464;
Best Local Similarity 23.7%; Pred. No. 1.70e-01;
Matches 9; Conservative 15; Mismatches 12; Indels 2; Gaps 2;

Db 172 LEIRDTAI-EAIEFLYDEAHDKWIKNGNFRVYKLSR 208
|||||: |||||: |||||: |||||:
QY 1 LDLOEASVADKLSFGKMAETRGTTWTPHOGNN-HVRLR 37

RESULT 6
ID O76253 PRELIMINARY; PRT; 344 AA.
AC O76253;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
OS DROSOPHILA TEISSIERI (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-DTB24;
RX MEDLINE; 96270054.
RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RT "The mariner transposable element in natural populations of
RT Drosophila teissieri.";
RL J. MOL. EVOL. 42:669-675(1996).
RN [2]
RP SEQUENCE FROM N.A.

| Query Match | 29.1% | Score 77 | DB 5 | Length 345 |
|-----------------------|-------|--------------------|------|----------------------------------|
| Best Local Similarity | 28.9% | Pred. No. 2.53e-01 | | |
| Matches | 11 | Conservative | 14 | Mismatches 10; Indels 3; Gaps 3; |

```

RN 10
ID 061452 PRELIMINARY; PRT; 345 AA.
AC 061452;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
DS DROSOPHILA SIMULANS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 93106423.
RA CAPY P., DAVID J.R., HARTL D.L.;
RT "Evolution of the transposable element mariner in the Drosophila
  melanogaster species group.";
RL GENETICA 86:37-46(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 92201636.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Sequence analysis of active mariner elements in natural populations
  of Drosophila simulans.";
RL GENETICS 130:499-506(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 96270054.
RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RT "The mariner transposable element in natural populations of
  Drosophila teissleri.";
RL J. MOL. EVOL. 42:669-675(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RA CAPY P.;
RT SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RL EMBL; AF037056; G3136134; -.
SQ SEQUENCE 345 AA; 40789 MW; EFCFDF0F CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
Best Local Similarity 28.9%; Pred. No. 2.53e-01;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSGQAVSNRLREMGI-QKVGWVPHLINER-QMER 131
|:::|:::| |:::| |:::| |:::|
QY 1 LDQEASVADKLS-FGKMAETRGTWTPHQGNHVR LPR 37

RESULT 11
ID 061455 PRELIMINARY; PRT; 345 AA.
AC 061455;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
DS DROSOPHILA SIMULANS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 93106423.
RA CAPY P., DAVID J.R., HARTL D.L.;
RT "Evolution of the transposable element mariner in the Drosophila
  melanogaster species group.";
RL GENETICA 86:37-46(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 92201636.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Sequence analysis of active mariner elements in natural populations
  of Drosophila simulans.";
RL GENETICS 130:499-506(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RX MEDLINE; 96270054.
RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RT "The mariner transposable element in natural populations of
  Drosophila teissleri.";
RL J. MOL. EVOL. 42:669-675(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-MADB;
RA CAPY P.;
RT SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RL EMBL; AF037056; G3136134; -.
SQ SEQUENCE 345 AA; 40789 MW; EFCFDF0F CRC32;

```

```

RP      SEQUENCE FROM N.A.
RC      TRANSPONON-MARINER-BORDA;
RX      MEDLINE; 93106423.
RA      CAPY P., DAVID J.R., HARTL D.L.;
RT      "Evolution of the transposable element mariner in the Drosophila
RL      melanogaster species group.";
RN      GENETICA 86:37-46(1992).
[2]
RP      SEQUENCE FROM N.A.
RC      TRANSPONON-MARINER-BORDA;
RX      MEDLINE; 92201636.
RA      CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT      "Sequence analysis of active mariner elements in natural populations
RL      of Drosophila simulans.";
RN      GENETICS 130:499-506(1992).
[3]
RP      SEQUENCE FROM N.A.
RC      TRANSPONON-MARINER-BORDA;
RA      CAPY P.;
RL      SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR      EMBL; AF037060; G3136142; -;
SQ      SEQUENCE 345 AA; 40774 MW; 065A5AAB CRC32;

Query Match          29.1%; Score 77; DB 5; Length 345;
Best Local Similarity 28.9%; Pred. No. 2.53e-01;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

DB      96 LEYSOQAVSNRLREMGKI-QKVGKRWVPHELNER-QMER 131
      1:::1:::1 11::11111::1
QY      1 LDIQASVADKLS-FGKMAETRGTWTPHOGNNHVRLLPR 37

```

```

RESULT 12
ID 062616 PRELIMINARY; PRT; 345 AA.
AC 062616;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
OC DROSOPHILA SIMULANS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
RX MEDLINE; 93106423.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Evolution of the transposable element mariner in the Drosophila
RL melanogaster species group.";
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
RX MEDLINE; 92201636.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Sequence analysis of active mariner elements in natural populations
RL of Drosophila simulans.";
RN [3]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
RA CAPY P.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF037057; G3136136; -
DR EMBL; AF037052; G3136126; -
DR EMBL; AF037053; G3136128; -
SQ SEQUENCE 345 AA; 40798 MW; DE15C886 CRC32;

```


D**b** 96 LEVSGQAVSNRLREMGKI-QKVGWRWPHETNER-QMER 131
| : : : | | : | | | : : :
QY 1 LDLOEASVDKLKLS-FGKMAETRGTWTWPHCGNNHVRLLPR 37

```

RESULT 13
ID 061454 PRELIMINARY; PRT; 345 AA.
AC 061454;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSPOSASE.
OC DROSOPHILA SIMULANS (FRUIT FLY).
OC EURARCYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-PRA;
RX MEDLINE; 93106423.
RA CAPY P., DAVID J.R., HARTL D.L.;
RT "Evolution of the transposable element mariner in the Drosophila
RL melanogaster species group.";
RL GENETICA 86:37-46(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-PRA;
RX MEDLINE; 92201636.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Sequence analysis of active mariner elements in natural populations
RL of Drosophila simulans.";
RL GENETICS 130:499-506(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-PRA;
RA CAPY P.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF037059; G3136140; -.
SO SEQUENCE 345 AA; 40807 MW; B94426B0 CRC32;

```

| | | | | |
|--------------------------|--------|--------------------------------------|-------|-------------|
| Query Match | 29.18; | Score 77; | DB 5; | Length 345; |
| Best Local Similarity | 28.98; | Pred. No. 2.53e-01; | | |
| Matches 11; Conservative | | 14; Mismatches 10; Indels 3; Gaps 3; | | |

Db 96 LEVSGQAVSNRLREMGRI-QKVGRAWPHELNER-QMER 131
 1:::1::1 11:: 11111:::1
 QY 1 LDLOEASVADKLS-FGKMAETRGTWIPHOGNHHVRLPR 37

| RESULT | 14 | | |
|--------|---|---|--------------|
| ID | 061446 | PRELIMINARY; | PRT; 345 AA. |
| AC | 061446; | | |
| DT | 01-AUG-1998 | (TREMBLREL. 07, CREATED) | |
| DT | 01-AUG-1998 | (TREMBLREL. 07, LAST SEQUENCE UPDATE) | |
| DT | 01-NOV-1998 | (TREMBLREL. 08, LAST ANNOTATION UPDATE) | |
| DE | TRANSPOSASE. | | |
| OS | DROSOPHILA SECHELLIA (FRUIT FLY). | | |
| OC | EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA; | | |
| OC | PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA; | | |
| OC | DROSOPHILIDAE; DROSOPHILA. | | |
| RN | (1) | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TRANSPOSON-MARINER-PA2SECH; | | |
| RX | MEDLINE; 92071978. | | |
| RA | CAPY P., MARUYAMA K., DAVID J.R., HARTL D.L.; | | |
| RT | "insertion sites of the transposable element mariner are fixed in the | | |
| RT | genome of Drosophila sechellia."; | | |
| RL | J. MOL. EVOL. 33:450-456(1991). | | |
| RN | (2) | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TRANSPOSON-MARINER-PA2SECH; | | |
| RX | MEDLINE; 93106423. | | |
| RA | CAPY P., DAVID J.R., HARTL D.L.; | | |

RT "Evolution of the transposable element mariner in the *Drosophila*
RT *melanogaster* species group.";
RL GENETICA 86:37-46(1992).

RP .SEQUENCE FROM N.A.
RC TRANSPON-MARINER-PA2SECH;
RX MEDLINE; 96270054.
RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RT "The mariner transposable element in natural populations of
RT *Drosophila teissieri*.";
RL J. MOL. EVOL. 42:669-675(1996).
RN [4]
RP .SEQUENCE FROM N.A.
RC TRANSPON-MARINER-PA2SECH;
RA CAPY P.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF035568; G313611; -.
SQ .SEQUENCE 345 AA; 40768 MW; 6C1D83C8 CRC32;

| | | | | |
|--------------------------|--------|---------------------|-----------|-------------|
| Query Match | 29.1%; | Score 77; | DB 5; | Length 345; |
| Best Local Similarity | 28.9%; | Pred. No. 2.53e-01; | | |
| Matches 11; Conservative | | 14; Mismatches 10; | Indels 3; | Gaps 3; |

Db 96 LEVSAQAVSNRLREMGRI-QKVGRAWDPHELNER-QMER 131
 1 1 : : : : : 1 1 : : : : : 1
 QY 1 LDLDQESVADKLS-FGKMAETRGITWTPHQGNHHVRLPR 37

| | | | |
|--------|--|---|--------------|
| RESULT | 15 | | |
| ID | 061450 | PRELIMINARY; | PRT; 345 AA. |
| AC | 061450; | | |
| DT | 01-AUG-1998 | (TREMBLREL. 07, CREATED) | |
| DT | 01-AUG-1998 | (TREMBLREL. 07, LAST SEQUENCE UPDATE) | |
| DT | 01-NOV-1998 | (TREMBLREL. 08, LAST ANNOTATION UPDATE) | |
| DE | TRANSPOSASE. | | |
| OS | DROSOPHILA SIMULANS (FRUIT FLY). | | |
| OC | EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA; | | |
| OC | PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA; | | |
| OC | DROSOPHILIDAE; DROSOPHILA. | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TRANSPOSON-MARINER-L8; | | |
| RX | MEDLINE; 93106423. | | |
| RA | CAPY P., DAVID J.R., HARTL D.L.; | | |
| RT | "Evolution of the transposable element mariner in the Drosophila | | |
| RT | melanogaster species group."; | | |
| RL | GENETICA 86:37-46(1992). | | |

```

RP SEQUENCE FROM N.A.
RC TRANSPOSON-MARINER-L8;
RX MEDLINE; 92201636.
RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
RT "Sequence analysis of active mariner elements in natural
RL of Drosophila simulans.";
RN GENETICS 130:499-506(1992).
RP [3]
RC SEQUENCE FROM N.A.
RX TRANSPOSON-MARINER-L8;
RA MEDLINE; 96270054.
RT BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RL "The mariner transposable element in natural populations of
RN Drosophila telstieri.";
RP J. MOL. EVOL. 42:669-675(1996).
RC [4]
RX SEQUENCE FROM N.A.
RA TRANSPOSON-MARINER-L8;
RT CAPY P.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN EMBL; AF037054; G3136130; -.
SQ SEQUENCE 345 AA; 40788 MW; 435FE906 CRC32;

```


Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;
Db 96 LEVSGQAVSNRLREMGKI-QKVGWVPHELNER-QMER 131
|: : : |: : | | | | |: : : |
QY 1 LDLQEASVADKLS-FGKMAETRGTTWPHQGNHVRLLPR 37

Search completed: Wed Sep 29 14:08:46 1999
Job time : 16 secs.

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA

FEATURES
source location/Qualifiers
1..1347
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19q13"
1..765
 /note="androgen regulated"
 /codon_start=1
 /product="serine protease prostate"
 /protein_id="AAD21580.1"
 /db_xref="PID:94512030"
 /db_xref="GI:4512030"
 /translation="MATAGNPMGWFLGYLLGVAGSLVSGSCQIINGEDCSPHSOPW
OALVMENELFCSGVLVHPQWLSAHCNOSTYIGLGLHLEADQEPGSOMEASLS
VRHPEYNRPILLANDMLIKLDESSESPTIRISIASOCPTAGNSCLVSGWGLLANGR
MPTVLQCVANSVSEEVCSKLYDPLIHPSMFCAGGGHDKDSNGDSGDPICNGYLQ
GLVSFGKAPCGQVGVPGVYTNLCKFTIEWIENTIVQAS"
1327..1332

polyA_site
BASE COUNT 269 a 489 c 334 g 255 t
ORIGIN

Query Match 32.2%; Score 360.8; DB 42; Length 1347;
Best Local Similarity 96.8%; Pred. No. 1.2e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 2 CGCACTCGCAGCCCTGGCAGCGCGGCACACTGTGTCATGGAAACGAATTGTTCTGCTCGGGCG 61
|||||
Db 116 CGCACTCGCAGCCCTGGCAGCGCGGCACACTGTGTCATGGAAACGAATTGTTCTGCTCGGGCG 175

OY 62 TCCTGTGCATCCGCCAGTGGTGCTGTGCAGCCGCACACACTGTTCCAGAAGTCTCTACACCA 121
|||||
Db 176 TCCTGTGCATCCGCCAGTGGTGCTGTGCAGCCGCACACACTGTTCCAGAAGTCTCTACACCA 235

OY 122 TCGGGCTGGCCTGCACAGTCTTGAGGCGCCGACCAGGAGCCAGGAGCCAGATGTTGAGG 181
|||||
Db 236 TCGGGCTGGCCTGCACAGTCTTGAGGCGCCGACCAGGAGCCAGGAGCCAGATGTTGAGG 295

OY 182 CCAGCCTCTCCGTACGGCACCCAGAGTAGACACAGACCCTTGTCTCGCTAACGACCTCATGC 241
|||||
Db 296 CCAGCCTCTCCGTACGGCACCCAGAGTAGACACAGACCCTTGTCTCGCTAACGACCTCATGC 355

OY 242 TCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTTGCTT 301
|||||
Db 356 TCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTTGCTT 415

OY 302 CGCAGTGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGC 361
|||||
Db 416 CGCAGTGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGC 475

OY 362 ATGCTGTGATTGCCATCCAG 381
|| | | | | | |
Db 476 GCAGAAATGCCTACCGTGTG 495

RESULT 2
SSU76256 1140 bp mRNA MAM 03-JAN-1998
LOCUS
DEFINITION Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,
complete cds.
ACCESSION U76256
NID 92737920
VERSION U76256.1 GI:2737920
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 1140)
Slimmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,

| QY | DB | Query Match | Best Local Similarity | Score | DB 3; | Length | DB 3; | Length |
|----|-----|---|-----------------------|--------------------|----------------|-----------|---------|--------|
| QY | 182 | CCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCTTGTCTCGCTAACGACCTCATGC | 24.2%; | 270.6; | DB 3; | 1140; | | |
| DB | 304 | TCGGGCTGGGTCTGTGAGCCCGAAGAACCCCGCAGCCAGATGATGAGG | 84.9%; | Pred. No. 3.6e-66; | | | | |
| DB | 244 | TCCTGTGTCATCCGCAATGGGTGCTGTGTACGCCGACACTGCTTCCAAAATTCCTACACCA | Matches 303; | Conservative 0; | Mismatches 54; | Indels 0; | Gaps 0; | |
| QY | 184 | CGCACTCGCAGCCCTGCGAGCGCGGCACTGTGTCATGGAAGCAATTTGCTGCTCGGCGG | | | | | | |
| QY | 62 | TCCTGTGTCATCCGCAATGGGTGCTGTGTACGCCGACACTGTTTCCAGAACTCCTACACCA | | | | | | |
| QY | 122 | TCGGGCTGGGTCTGTGAGCCCGAAGAACCCCGCAGCCAGATGATGAGG | | | | | | |
| QY | 304 | TCGGGCTGGGTCTGTGAGCCCGAAGAACCCCGCAGCCAGATGATGAGG | | | | | | |
| QY | 182 | CCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCTTGTCTCGCTAACGACCTCATGC | | | | | | |

Db 364 CCAGCCTCTCCATTCAGCACCCAGATAACAGAGCCGTCTATGCCCAAGACCTCATGC 423

OY 242 TCATCAGTTGAGCAATCCGTGTCCAGATTCACACCATCCGAGCATCAGCATTCGCTT 301

Db 424 TCATCAAGTTGAAGAATCGGTGTCTGTGACACCGTCCGAAACATCAGCGTCTCT 483

OY 302 CGCAGTGCCTACCGGGGGAACCTCTTGCCTTCCTTCTGCTGGGTCTGCTGGCGA 358

Db 484 CCCAGTGCCTACCGGGGGAACCTCTTGCCTTCCTTCTGCTGGGTCTGCTGGCGA 540

RESULT 3

AF113141 5900 bp DNA PRI 25-MAR-1999

LOCUS Homo sapiens serine protease prostate gene, complete cds.

DEFINITION AF113141

ACCESSION AF113141

NID 94512031

VERSION AF113141.1 GI:4512031

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.

TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)

MEDLINE 99179024

REFERENCE 2 (bases 1 to 5900)

AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-1998) Chitrosience R & D, 1631 220th St SE, Bothell, WA 98021, USA

FEATURES

source

1. 5900

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13"

join(<1411..1471,2735..2897,3319..3569,3653..3789,5062..5817)

/product="serine protease prostate"

1411..1471

/number=1

join(1411..1471,2735..2897,3319..3569,3653..3789,5062..5817)

/note="androgen regulated"

/codon_start=1

/product="serine protease prostate"

/protein_id="AAD21581.1"

/db_xref="PID:94512032"

/db_xref="GI:4512032"

/translation="MATAGNPMWFLGLILGVAGSLVSGSCQIINGEDCSPHSQPM QAAIVMENELFCSGVLVHPQWVLSAHCFOFNSYTGILHSLLEADQEPGOMVEASLS VRHPEYNRPPLANDLMLIKDESSESPTIRSIASQCPAGNSCLVSGWGLLANGR MPVLIQCVNVSVEEYCSKLYDPLVHPMFCAAGGHQDKSCNGDSGGLICNGYLO GLVSFGKAPCGQGVGPVYTNLCKFTWIEKTVQAS"

2735..2897

/number=2

3319..3569

/number=3

3653..3789

/number=4

5062..5817

/number=5

BASE COUNT 1417 a 1643 c 1569 g 1271 t

ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 5900;

Best Local Similarity 98.8%; Pred. No. 1.1e-61;

Matches 257; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 105 CCAGACTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGACCCAGG 164

Db 3313 CCACAGCTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGACCCAGG 3372

OY 165 GAGCCAGATGGTGAGGGCCAGCCCTCTCCGTACAGGCACCCAGAGTACACAGACCCCTGCT 224

Db 3373 GAGCCAGATGGTGAGGGCCAGCCCTCTCCGTACAGGCACCCAGAGTACACAGACCCCTGCT 3432

OY 225 CGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTCCGAGTCTGACACCATCCG 284

Db 3433 CGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTCCGAGTCTGACACCATCCG 3492

OY 285 GAGCATCAGCATTCGCTTCGAGTGCCTACCGCGGGGAACCTTGCCTCGTTTCGCTG 344

Db 3493 GAGCATCAGCATTCGCTTCGAGTGCCTACCGCGGGGAACCTTGCCTCGTTTCGCTG 3552

OY 345 GGGTCTGCTGGCGAACGATG 364

Db 3553 GGGTCTGCTGGCGAACGATG 3572

RESULT 4

AF135023 4740 bp DNA PRI 20-APR-1999

LOCUS Homo sapiens kallikrein-like protein 1 KLK-L1 gene, partial cds.

DEFINITION AF135023

ACCESSION AF135023

NID 94589272

VERSION AF135023.1 GI:4589272

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.

TITLE Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4

JOURNAL Blochem. Biophys. Res. Commun. (1999) In press

REFERENCE 2 (bases 1 to 4740)

AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.

TITLE Direct Submission

JOURNAL Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

FEATURES

source

1. 4740

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13.3-q13.4"

join(<2263..2425,2847..3097,3181..3317,4588..>4740)

/product="kallikrein-like protein 1 KLK-L1"

join(<2263..2425,2847..3097,3181..3317,4588..4740)

/codon_start=3

/product="kallikrein-like protein 1 KLK-L1"

/protein_id="AAD26424.1"

/db_xref="PID:94589273"

/db_xref="GI:4589273"

/translation="SLVSGSCQIINGEDCSPHSQPMQAAIVMENELFCSGVLVHPQW VLSAHCFOFNSYTGILHSLLEADQEPGOMVEASLSVRHPEYNRPPLANDLMLIKLD ESVSESDTIRSIASQCPAGNSCLVSGWGLLANGRMPVLIQCVNVSVEEYCSKLYDPLVHPMFCAAGGHQDKSCNGDSGGLICNGYLOGLVSFGKAPCGQGVGPVYTN LCKFTWIEKTVQAS"

BASE COUNT 1106 a 1275 c 1308 g 1051 t

ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 4740;

Best Local Similarity 98.8%; Pred. No. 1e-61;


```

mat_peptide      /note="encodes propeptide"
                  145. .816
BASE COUNT      289 a      386 c      286 g      276 t
ORIGIN

```

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 19.0%; | Score 212.8; | DB 12; | Length 1237; |
| Best Local Similarity | 74.4%; | Pred. No. 7.9e-50; | | |
| Matches 268; Conservative | 0; | Mismatches 92; | Indels 0; | Gaps 0; |

Qy 2 CGCACTCGAGCCCTGGCAGGCGGCACTGGTCATGAAACGAATTGTCTGCTCGGGC 61
170 CACACTCGCAACCCCTGGCAGGCGGCACTGTTCACGAAGACGGTTTTTCTGCTCGGGAG 229

| | | | |
|----|-----|--|-----|
| QY | 62 | TCCCTGGTCATCCGAGTGGGTGCTGTACGCCGCACACTGTTTCCAGAAGACTCCTACACA | 121 |
| | | | |
| | | | |
| Db | 230 | TCTTGTCATCCCGAGTGGGTGCTGTACGAGCACATTGCTTACAGGAGTCTCAATCG | 289 |

QY 122 TCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAGAGCCAGCGAACCATGTGTGAGC 181
- - - - -
Db 290 TGGACTGGGCGCTGCATTAACCTGAAAGGCTGCCAAGAGCGCTGGCAGCCGGATGTTAGGG 349

QY 182 CCAGCCTCTCCGTACGGCACCCAGAGTACAACGACCCTTGCTCGCTAACGACCTCATGC 241
|| ||||||| | ||||||| | | |||| | | ||||||| |||||||
Db 350 CCCACCTCTTCATCCAGCACCCCAACTTCACAGATCTCTTTTGAAGAAGATCTCATGC 409

QY 242 TCATCAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGACATCAGATTGCTT 301
||||| | |||| | ||| |||||| | || | |||||
Db 410 TCATCAACTGACGAGTCAAGTATAGAGCTTAACACTATCAGAGCATCCCTGTGGCTA 469

QY 302 CGCAGTGCCTACCGCGGGGAAGCTTTGCTTCGTGGCTGGGGTCTGCTGGCGAAGC 361
| | | | | | | | | | | | | | | | | | | | |
Db 470 CCCAATGCCCACTCTTGAGATACCTGCTAGTCTCTGGTGGGGTCACTAAAGAATG 529

| | | | |
|------------|---|-----------|-----------------|
| RESULT | 7 | | |
| HS48G12/c | | | |
| LOCUS | HS48G12 | 199016 bp | DNA |
| DEFINITION | Human DNA sequence from clone 48G12 on chromosome Xq27.1-27.3. Contains 5TSS and 5SSS, complete sequence. | | |
| | | | PRI 08-OCT-1998 |

| | |
|-----------|---|
| ACCESSION | AL031054 |
| NID | 93550059 |
| VERSION | AL031054.1 |
| KEYWORDS | GI:3550059 |
| SOURCE | HTG. |
| ORGANISM | human. Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 199016) |

| | |
|----------------|---|
| AUTHORS | Pearce, A. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (21-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk |
| COMMENT | On Sep 8, 1998 this sequence version replaced q1:3451161. |

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 48G12. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrx>

48G12 is from the library RPC11 constructed at the Roswell park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:PCYPAC2>.

location/Qualifiers
1. .199016

FEATURES
source

```

repeat_region
/note="MLT2_internal repeat: matches 4550. .5668 of
consensus"
11. .1153
repeat_region
/note="MLT2CA repeat: matches 1. .300 of consensus"
1155. .1450
repeat_region
/note="MLT2CB repeat: matches 316. .444 of consensus"
1425. .1544
repeat_region
/note="MLT2CB repeat: matches 316. .444 of consensus"
1736. .1779
repeat_region
/note="22 copies 2 mer at 93% conserved"
1819. .1884
repeat_region
/note="MLT1D repeat: matches 288. .353 of consensus"
1914. .1943
repeat_region
/note="15 copies 2 mer at 100% conserved"
2061. .2100
repeat_region
/note="20 copies 2 mer at 83% conserved"
2138. .2173
repeat_region
/note="MLT2FA repeat: matches 416. .450 of consensus"
complement(2636. .2933)
repeat_region
/note="AluSg repeat: matches 299. .2 of consensus"
complement(3141. .3601)
repeat_region
/note="MLT1D repeat: matches 482. .1 of consensus"
3857. .4383
misc_feature
/note="match: GSS B59439"
3901. .3962
repeat_region
/note="FLAM repeat: matches 71. .132 of consensus"
3967. .4000
repeat_region
/note="17 copies 2 mer aa 82% conserved"
complement(4206. .4385)
repeat_region
/note="L1PA4 repeat: matches 893. .714 of consensus"
5064. .5354
repeat_region
/note="L1MA7 repeat: matches 767. .1036 of consensus"
complement(7448. .8028)
repeat_region
/note="L1MC1 repeat: matches 1074. .482 of consensus"
complement(8178. .8370)
repeat_region
/note="L1MC1 repeat: matches 491. .298 of consensus"
complement(8371. .9258)
repeat_region
/note="L1PA2 repeat: matches 893. .1 of consensus"
complement(9112. .11605)
repeat_region
/note="L1 repeat: matches 5390. .2936 of consensus"
complement(11602. .11892)
repeat_region
/note="L1MC1 repeat: matches 301. .5 of consensus"
complement(11750. .12175)
repeat_region
/note="L1 repeat: matches 5390. .4955 of consensus"
12180. .12479
repeat_region
/note="AluY repeat: matches 1. .299 of consensus"
complement(12527. .13544)
repeat_region
/note="L1 repeat: matches 4944. .3919 of consensus"
complement(13756. .13969)
repeat_region
/note="MIR repeat: matches 249. .30 of consensus"
complement(14025. .14226)
repeat_region
/note="MIR repeat: matches 262. .52 of consensus"
15048. .15735
repeat_region
/note="L1 repeat: matches 3616. .4317 of consensus"
16545. .17275
repeat_region
/note="L1ME2 repeat: matches 150. .911 of consensus"
complement(19351. .19472)
repeat_region
/note="MIR repeat: matches 253. .124 of consensus"
complement(19718. .19889)
repeat_region
/note="L1 repeat: matches 5189. .5008 of consensus"
19890. .20010
repeat_region
/note="AluJo repeat: matches 1. .124 of consensus"
complement(20041. .20304)

```


/note="L1 repeat: matches 5015. .4754 of consensus"
repeat_region complement(20316. .20556)
/note="L1MA4A repeat: matches 1047. .801 of consensus"
repeat_region complement(20602. .20758)
/note="L1 repeat: matches 2529. .2371 of consensus"
repeat_region complement(20749. .21857)
/note="L1 repeat: matches 2203. .1103 of consensus"
repeat_region complement(21324. .22281)
/note="MER25 repeat: matches 2136. .1177 of consensus"
repeat_region complement(22277. .22722)
/note="MER25 repeat: matches 1031. .582 of consensus"
repeat_region complement(22807. .22998)
/note="MER25 repeat: matches 205. .13 of consensus"
repeat_region 23002. .23044
/note="MSTA repeat: matches 384. .426 of consensus"
repeat_region 23045. .23443
/note="MST-INTERNAL repeat: matches 1. .394 of consensus"
repeat_region complement(23519. .23679)
/note="L1MB1 repeat: matches 914. .757 of consensus"
misc_feature 23755. .23876
/note="match: STS L41276"
repeat_region 23842. .24139
/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region complement(24613. .24899)
/note="AluJb repeat: matches 293. .7 of consensus"
repeat_region complement(25248. .25604)
/note="THE1B repeat: matches 364. .1 of consensus"
repeat_region complement(25605. .26715)
/note="THE1B-INTERNAL repeat: matches 1580. .471 of consensus"
repeat_region complement(26728. .26769)
/note="THE1C repeat: matches 44. .3 of consensus"
repeat_region complement(27295. .27591)
/note="AluSq repeat: matches 299. .1 of consensus"
repeat_region complement(28402. .28571)
/note="MER5B repeat: matches 174. .8 of consensus"
repeat_region 28721. .28894
/note="MIR repeat: matches 82. .258 of consensus"
repeat_region 29430. .30796
/note="L1 repeat: matches 4045. .5390 of consensus"
repeat_region 30659. .31555
/note="L1PA16 repeat: matches 1. .904 of consensus"
repeat_region 31625. .31648
/note="12 copies 2 mer tt 100% conserved"
repeat_region complement(31650. .31950)
/note="AluYa5 repeat: matches 301. .1 of consensus"
repeat_region complement(31962. .32114)
/note="L1MD1 repeat: matches 971. .806 of consensus"
repeat_region complement(31972. .32025)
/note="MER42C repeat: matches 1538. .1485 of consensus"
repeat_region complement(32061. .33135)
/note="MER42c repeat: matches 1301. .173 of consensus"
repeat_region complement(33214. .33525)
/note="L1MC3 repeat: matches 1362. .1054 of consensus"
repeat_region complement(35486. .36087)
/note="L1MA8 repeat: matches 617. .3 of consensus"
repeat_region complement(35931. .36173)
/note="L1 repeat: matches 5390. .5158 of consensus"
repeat_region 36176. .36573
/note="L1MA9 repeat: matches 620. .1016 of consensus"
repeat_region complement(38289. .38580)
/note="AluY repeat: matches 292. .1 of consensus"
repeat_region complement(38584. .40069)
/note="L1 repeat: matches 4630. .3154 of consensus"
repeat_region complement(40030. .40877)
/note="L1 repeat: matches 3086. .2235 of consensus"
repeat_region complement(40843. .40929)
/note="L1MA3 repeat: matches 148. .61 of consensus"
repeat_region complement(40909. .41341)
/note="L1 repeat: matches 5287. .4843 of consensus"
repeat_region 44097. .44291
/note="MIR repeat: matches 1. .204 of consensus"
repeat_region complement(44511. .44801)

/note="AluSx repeat: matches 285. .1 of consensus"
repeat_region 45620. .45721
/note="MIR repeat: matches 81. .189 of consensus"
repeat_region complement(45926. .46063)
/note="MLT1F repeat: matches 541. .379 of consensus"
repeat_region complement(47019. .47161)
/note="MIR repeat: matches 262. .120 of consensus"
repeat_region 47339. .47412
/note="MIR repeat: matches 64. .137 of consensus"
repeat_region 47884. .48136
/note="L1PA4 repeat: matches 643. .891 of consensus"
repeat_region complement(48404. .48890)
/note="L1PA3 repeat: matches 893. .406 of consensus"
repeat_region 49119. .49216
/note="MIR repeat: matches 146. .246 of consensus"
misc_feature complement(50414. .50631)
/note="match: STS G03807"
repeat_region complement(51021. .51433)
/note="MLT1D repeat: matches 505. .101 of consensus"

Query Match 12.4%; Score 138.4; DB 9; Length 199016;
Best Local Similarity 62.2%; Pred. No. 2.1e-28;
Matches 270; Conservative 0; Mismatches 156; Indels 8; Gaps 3;

QY 505 ACTGCATCACCCGGAACACTGTGATCACTAGCCAGCACCATAGTCTCCGAAGTCAGAC 564
Db 136146 ATTTCACCTCTAGGAACACTGTAAATTAATTAACCAAGCCAGAGCTCTACCTGATCAGAC 136087
QY 565 TATCATGATTACTGTGTGCTACTGTGCTGTCTATTGTACTAACCATGCCGATGTTAGGTG 624
Db 136086 TATTCGATTGCTGCTTGGCCCTGCTGCTATTATGATGCTAGCGCCACCCTGCTTTT 136027
QY 625 AA-----ATTAGGCTCACTTGGCCCTCAACCATCTTGGTATCCAGTTATCCCTCACTGAAT 678
Db 136026 GACAGTTGAGTGTGCAACTTGGCCACTGCCACTTGGGATCCAAATCATTCCCATGTGAT 135967
QY 679 TGAGATTTCCTGCTTCAGTGTGACGCAATCCACATAAATTCTGACCTACAGAGTGAGG 738
Db 135966 TTAATTGAGTTCAGTGCAGTCCCATGTTAGATCTGACATACAGAGAGAGC 135907
QY 739 GATCATATAGCTCTCAAGATGCTGTACTCCCTCACAAATTCATTCTCTCTGTGTA 798
Db 135906 AATTACAGGCTCTCAAGATGCGAGTGTGCCCTCAAAATCTATTTCAGAGAGTAT 135847
QY 799 GTGAAGGTGCGCCCTCTGAGACCTCCAGGGTGGGTGTGACAGTTC-ACAATGATGAATG 857
Db 135846 GGTCAAGGTATATCTTCTGACCCTTCACGCTGGGATGAATAGGTCTAAAGTGACTAATG 135787
QY 858 TATGATCGTGTCCCATTTACCAAGACCTTAAATCCC-TCATGCTCAGTACACCAAGGCC 916
Db 135786 TATTCACCACCAATCTCCCTAAGCCTTTGGATCCCTTCCTTTACATAAACCAAGGG 135727
QY 917 AGGTCTAGCATTTTC 930
Db 135726 AGATCAGGCATTTTC 135713

RESULT 8
HSU76377
LOCUS HSU76377 42179 bp DNA PRI 02-JAN-1999
DEFINITION Human olfactory receptor 17-93 (OR17-93) and olfactory receptor
17-201 (OR17-201) genes, complete cds.
ACCESSION U76377
NID 94098234
VERSION U76377.1 GI:4098234
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 42179)
AUTHORS Ferraz,C. and Demaille,J.G.
TITLE Sequence of cosmid ICRF105ch07155 of the olfactory receptor gene

| FEATURES | source |
|---|---|
| JOURNAL | cluster of human chromosome 17p13.3 |
| REFERENCE | unpublished (1996) |
| AUTHORS | 2 (bases 1 to 42179) |
| TITLE | Ferraz, C. and Demaille, J.G. |
| JOURNAL | Direct Submission |
| | Submitted (26-OCT-1996) CRBM, CNRS, 1919, route de Mende, PO Box 5051, Montpellier 34033, France |
| | Location/Qualifiers |
| | 1. .42179 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /clone="cosmid ICRF105CH07155" |
| | /chromosome="17" |
| | /map="17p13.3" |
| | 5285. .6250 |
| | /gene="OR17-201" |
| | 5285. .6250 |
| | /gene="OR17-201" |
| | /codon_start=1 |
| | /product="olfactory receptor 17-201" |
| | /protein_id="AAD00250.1" |
| | /db_xref="PID:94098236" |
| | /db_xref="GI:4098236" |
| | /translation="MSLQKIMEPEAGNRTAVAEFLLGLVQTEEMOPVVEVLLLFAY LVTIGNSILAAVLVEPKLHAPMFEFLGNLSVLDVGCITVTVPAMIGLLSHKSTIS YDACLSQLFEFFHLLAGMDCFLITAMAYDRLLAICQPLTYSTRMSQTVQRMVAAFWAC AFTNALHTVAMSTLNEFCGPNENHFECDLPQLSCSSSTOLNELLEFVAAAFMAVA PLFVSVSYAHVVAVLQIRSAEGRKAFSTCGSHLTVGIFGTGVSFYMRLGSVES SDKDKGVEMTVINPMLNPLIYSLRNTDVGALCOLLVERSLT" |
| | complement(17586. .18257) |
| | /gene="OR17-93" |
| | complement(17586. .18257) |
| | /gene="OR17-93" |
| | /codon_start=1 |
| | /product="olfactory receptor 17-93" |
| | /protein_id="AAD00249.1" |
| | /db_xref="PID:94098235" |
| | /db_xref="GI:4098235" |
| | /translation="MYFLYFSDESFLVAMAYDRVVAICFPMYHAICFLHYTAI MSPMLCLSVVLSVLTTFHAMLHTLMAKLCFCADNVIPIHFCDSALLKLACSDTR VNEWVIFINGLILVIPLLIIGSVARISSILKVPSSKGICAFSTCGSHLSVSLF YGVIVGLYLCPSANSSTLKDTVMAMTYVTVPMLTPFIYSLRNRDMKALERVICRRK NPFL" |
| BASE COUNT | 13441 a 8694 c 7851 g 12193 t |
| ORIGIN | |
| Query Match | 12.0% Score 134.2; DB 11; Length 42179; |
| Best Local Similarity | 65.4% Pred. No. 2.4e-27; |
| Matches 291; Conservative | 0; Mismatches 143; Indels 11; Gaps 6; |
| QY 497 TACTGCTCACTGCATCACCAGCAACACTGTGATCAACTAGCCAGCACCATAGTTCT-CCG 555 | |
| Db 10998 TCCTGTCAAGTTTCACTCTAGGAACACCATGATTAATTAGCCAAATGCCAGAGCTCTACAC 11057 | |
| QY 556 AAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTCTATTGTACTAACCATGCCGAT 615 | |
| Db 11058 AAGTCAGACTATTCGATTGCTGTTTGCCTCTGCTGTCATTAATGCTGCTACAGCCAA 11117 | |
| QY 616 GTTTAGGTGAA-----ATTAGCGTCACTTGCCCTCAACCATCTTGGTATCCAGTTATCC 669 | |
| Db 11118 CTGTGCTTTAATGGTTAGTGTCTGCCACTTGGCCCTGCCACCTCGGATCCAAATTATTC 11177 | |
| QY 670 TCACTGAATGAGATTTCCTGCTTCAAGTGTGACGATGCCATCCAC-ATAATTCTGACCTAC 728 | |
| Db 11178 CCACTGCATTTAAATTATTAAGTTGAGTGACTGCAGTTCCTCCACTGTAAGATCTGACATAC 11237 | |
| QY 729 AGAGGTGAGGATCATATAGCTCTTCAAGGATGCTGTGACTCCCTCACAATTCATTTTC 788 | |
| Db 11238 AGAGAGAGCAATCATAGAGCTCTTGAAGATGACAGGTGCTGCCCTCACAATCTATTTC 11297 | |
| QY 789 TC-CTGTTAGTGAAGGTGCGCCCTCTGAGAGCTCCACAGGCTGGGTGTGACAGTC-AC 846 | |

| Db | Accession | Gene | Species | Location | Reference |
|------------|--|---|---------|----------|-------------|
| Db | 11298 | ACAAAGTATTAATCAAGGGGTATATCTTCTTGCGCCCTCCACGCTGGATGAGTAGCTCTAA | 11357 | | |
| QY | 847 | AATGATGAATGATGATCGTGTCCCATTTACCAAGCCTTAAAT-CCCTCATGCTCAG | 905 | | |
| Db | 11358 | AGTGACTAATCCACTCCAGCATCCCTATCTCCCTAAGCCTTTGGATTCCTTCCTACAT | 11417 | | |
| QY | 906 | TACACCAGGCGAGTCTAGCATTTC | 930 | | |
| Db | 11418 | TAAACCAAGAGATCAGGCATTTC | 11442 | | |
| RESULT | 9 | | | | |
| LOCUS | HSU78308/c | | | | |
| DEFINITION | HSU78308 | 39038 bp | DNA | PRI | 02-JAN-1999 |
| ACCESSION | NID | g4098463 | | | |
| VERSION | U78308 | | | | |
| KEYWORDS | U78308.1 | GI:4098463 | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | | |
| AUTHORS | Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | | |
| TITLE | 1 (bases 1 to 39038) | | | | |
| JOURNAL | Ferraz, C. and Demaille, J.G. | | | | |
| FEATURES | Sequence of cosmid ICRF105CE06173 of human chromosome 17p13.3 | | | | |
| SOURCE | olfactory receptor gene cluster, containing genes OR17-32 and | | | | |
| | OR17-201-1, and pseudogene OR17-01 | | | | |
| | Unpublished | | | | |
| | 2 (bases 1 to 39038) | | | | |
| | Ferraz, C. and Demaille, J.G. | | | | |
| | Direct Submission | | | | |
| | Submitted (15-NOV-1996) CRBM, CNRS, 1919, route de Mende, PO Box | | | | |
| | 5051, Montpellier 34033, France | | | | |
| | Location/Qualifiers | | | | |
| | 1. 39038 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /chromosome="17" | | | | |
| | /map="17p13.3" | | | | |
| | /clone="ICRF105CE06173" | | | | |
| | complement(9530..10495) | | | | |
| | /gene="OR17-201-1" | | | | |
| | complement(9530..10495) | | | | |
| | /gene="OR17-201-1" | | | | |
| | /codon_start=1 | | | | |
| | /product="olfactory receptor olfr17-201-1" | | | | |
| | /protein_id="AAD00277.1" | | | | |
| | /db_xref="PID:g4098465" | | | | |
| | /db_xref="GI:4098465" | | | | |
| | /translation="MSIQKLMEPEAGTNRATAVEFILLGLVQTEEMQPVFVLLLFAY | | | | |
| | LVTIGNLSTILAAVLVEPKLHAPMYEFLNLSVLDVGCITVYPAMLGRLLSHKSTIS | | | | |
| | YDACLSQLFFEFHLLAGMDCFLTLTAMAYDRLLAICQPLTYSTRSQTQRMVLAASWAC | | | | |
| | AFTNALHTHTVAMSTLNFCEGPNENHFCYCDLPOLFOLSCSSTOLNELLFVAAAFMAYA | | | | |
| | PLVIVSVYAHVAAVLAQIRSAEGRRKAFSTCGSHLTVGITGVFSYMRIGSVES | | | | |
| | SDKRGVGVEMTVINPMLNPLIYSLRNTDVQALCQLLVGERSLT" | | | | |
| | 26023..26967 | | | | |
| | /gene="OR17-32" | | | | |
| | 26023..26967 | | | | |
| | /gene="OR17-32" | | | | |
| | /codon_start=1 | | | | |
| | /product="olfactory receptor olfr17-32" | | | | |
| | /protein_id="AAD00276.1" | | | | |
| | /db_xref="PID:g4098464" | | | | |
| | /db_xref="GI:4098464" | | | | |
| | /translation="MMGQNCFTSIDFLLGLPIQPEOONLCYALFLAMYLTTLLGNLL | | | | |
| | ILVILRLDSHLHTPMYLFSLNLSFDLCFSSVTIPKLLQNMQNDPSIPYADCLTOMY | | | | |
| | FLLEGLDLESEFLVAMAYDRYVIAICPLHYTAIMSPMCLTIVALSMTTFHAMLHT | | | | |
| | LMARLCFCADNVIPHFCDMSALKLAFSDTRVNEWVIFIMGGLIVIPFLILGSY | | | | |
| | ARIVSSILKVPSSKGIKAFSTCGSHLSVSVLFGYTGIVGLYLCSSANSSTLKDTVMAM | | | | |
| | MTVVTPLNPLFIYSLRNDRMGALSRVTHOKTFFSL" | | | | |

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
TITLE Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 61450)
JOURNAL Phillips, S.
COMMENT Direct Submission
Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Dec 2, 1998 this sequence version replaced gi:3676174.
requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 15D7. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 34I8 (AL021918) is at 61351. In this
sequence. The true right end of clone 97D16 (AL009179) is at 100 in
this sequence. This sequence has been finished according to
sequence map criteria as follows. An attempt is made to resolve all
sequencing problems, such as compressions and repeats, but not
necessarily within known annotated human repeat sequence elements
(e.g. Alu). Where the sequence is ambiguous, there is an annotation
using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration by the Sanger
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,
David Ruddy, Jeffrey Gruen. Further information can be found at
<http://www.sanger.ac.uk/NGP/Chr6/>
15D7 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:pcvpac2>.
FEATURES
Location/Qualifiers
1..61450
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="15D7"
/map="p22.1-22.3"
/clone_lib="RPC11"
381..821
misc_feature
/note="match: GSS AQ213610"
643..950
repeat_region
/note="AluSp repeat: matches 1..309 of consensus"
1312..1385
repeat_region
/note="tRNA-Arg-CGC repeat: matches 1..74 of consensus"
1369..1486
misc_feature
/note="match: STS Z78001"
complement(1494..1797)
/note="match: STS G02930"
1960..2001
repeat_region
/note="21 copies 2 mer ag 76% conserved"
1968..2003
repeat_region
/note="9 copies 4 mer agac 92% conserved"
2803..3107
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
3292..3366
repeat_region
/note="tRNA-Ile-ATT repeat: matches 1..75 of consensus"
3673..3730
repeat_region
/note="MER6A repeat: matches 284..340 of consensus"
5345..5643
repeat_region
/note="AluX repeat: matches 10..310 of consensus"
5751..5899
repeat_region
/note="FLAM_A repeat: matches 1..125 of consensus"
5903..5954
repeat_region
/note="26 copies 2 mer tg 79% conserved"
5927..5954
repeat_region
/note="7 copies 4 mer tgtg 100% conserved"

misc_feature
6100..6268
/note="match: GSS B55592"
6208..6511
repeat_region
/note="AluSq repeat: matches 1..296 of consensus"
6564..6804
repeat_region
/note="Alu repeat: matches 1..301 of consensus"
7108..7183
repeat_region
/note="tRNA-Phe-TTC repeat: matches 1..73 of consensus"
7205..7250
repeat_region
/note="23 copies 2 mer tt 96% conserved"
7466..7791
repeat_region
/note="AluSp repeat: matches 1..313 of consensus"
8151..8461
repeat_region
/note="Alu repeat: matches 3..304 of consensus"
8462..8579
repeat_region
/note="MER92C repeat: matches 200..311 of consensus"
8580..8664
repeat_region
/note="MER6A repeat: matches 1..84 of consensus"
8785..8920
repeat_region
/note="MER92C repeat: matches 414..552 of consensus"
9471..9769
repeat_region
/note="AluX repeat: matches 1..300 of consensus"
9773..9804
repeat_region
/note="16 copies 2 mer tt 91% conserved"
9882..10170
repeat_region
/note="AluO repeat: matches 1..297 of consensus"
10186..10368
repeat_region
/note="L1MC2 repeat: matches 6158..6325 of consensus"
10370..10661
repeat_region
/note="AluSq repeat: matches 1..307 of consensus"
10676..10978
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
10991..11687
repeat_region
/note="L1MC2 repeat: matches 5173..5870 of consensus"
11682..12335
repeat_region
/note="L1MCA repeat: matches 663..1295 of consensus"
12236..12548
repeat_region
/note="AluX repeat: matches 1..311 of consensus"
12549..12862
repeat_region
/note="L1MCA repeat: matches 371..663 of consensus"
12937..13246
repeat_region
/note="Alu repeat: matches 1..309 of consensus"
13294..13389
repeat_region
/note="L1MD2 repeat: matches 6242..6330 of consensus"
13390..13687
repeat_region
/note="AluX repeat: matches 12..310 of consensus"
13688..13916
repeat_region
/note="L1MD2 repeat: matches 6031..6242 of consensus"
13917..14217
repeat_region
/note="AluX repeat: matches 1..303 of consensus"
14218..14516
repeat_region
/note="L1MD2 repeat: matches 5660..6031 of consensus"
14522..14691
repeat_region
/note="L1M4 repeat: matches 2357..2493 of consensus"
14692..15025
repeat_region
/note="AluX repeat: matches 1..310 of consensus"
15026..15055
repeat_region
/note="L1M4 repeat: matches 2324..2357 of consensus"
15158..15588
repeat_region
/note="L1MB4 repeat: matches 5725..6183 of consensus"
15605..15784
repeat_region
/note="AluDb repeat: matches 1..271 of consensus"
15785..15850
repeat_region
/note="L1MB4 repeat: matches 5667..5733 of consensus"
15851..16168
repeat_region
/note="AluSc repeat: matches 1..306 of consensus"
16169..16739
repeat_region
/note="L1MB4 repeat: matches 5176..5667 of consensus"
16740..17052
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
17053..17519
repeat_region
/note="L1MB4 repeat: matches 4696..5176 of consensus"
17541..19073
repeat_region

[illegible]

| Db | Accession | Gene | Length | Score | DB | Length | Score | DB |
|---------------------------|--|---|--------|-------|-------------|--------|-------|----|
| Db | 36995 | TGCTCAAGGGTGGATCTTCTGGACCTTCCAGAGCTGGAGTAGGTCTAAAGTGAATAA | 36936 | | | | | |
| QY | 856 | TGATATGATCGTGTCCCATTTACCCAAAGCCTTTAAATCCCTCATGCT-CAGTACACCAGG | 914 | | | | | |
| Db | 36935 | TTCACTCCACCATTCCCAATCTCCCTAAGCCTTTGGATCCCTTCTCTACATTAATCAAG | 36876 | | | | | |
| QY | 915 | GCAGTCTAGCATTTTC | 930 | | | | | |
| Db | 36875 | AGAGATCAGGCATTTTC | 36860 | | | | | |
| RESULT | 12 | | | | | | | |
| LOCUS | AF070718 | 166832 bp | DNA | PRI | 02-JUL-1998 | | | |
| DEFINITION | Homo sapiens BAC clone 529F11 from 8q21, complete sequence. | | | | | | | |
| ACCESSION | AF070718 | | | | | | | |
| NID | 93283634 | | | | | | | |
| VERSION | AF070718.1 | GI:3283634 | | | | | | |
| KEYWORDS | | | | | | | | |
| SOURCE | human. | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | | | | |
| AUTHORS | 1 (bases 1 to 166832) | | | | | | | |
| TITLE | Tauchi, H., Matsura, S., Isomura, M., Komatsu, K. and Nakamura, Y. | | | | | | | |
| JOURNAL | Direct Submission | | | | | | | |
| COMMENT | Submitted (03-JUN-1998) Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan | | | | | | | |
| | Mapping and sequence information: | | | | | | | |
| | This chromosome 8 clone was provided by S. Matsura, H. Tauchi, and K. Komatsu at Department of Radiation Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan. | | | | | | | |
| | Sequencing was performed by H. Tauchi, M. Isomura and Y. Nakamura at Human Genome Center, Institute of Medical Science, The University of Tokyo. | | | | | | | |
| FEATURES | | | | | | | | |
| source | Source information: | | | | | | | |
| | Clone 529F11 is isolated from the human BAC library available from Research Genetics, Inc. The library contains cloned DNA from the male fibroblast cell line 978SK. | | | | | | | |
| | Vector: pBelOBAC11. | | | | | | | |
| | Location/Qualifiers | | | | | | | |
| | 1.166832 | | | | | | | |
| | /organism="Homo sapiens" | | | | | | | |
| | /db_xref="taxon:9606" | | | | | | | |
| | /chromosome="8" | | | | | | | |
| | /map="8q21" | | | | | | | |
| | /clone="529F11" | | | | | | | |
| | /sex="male" | | | | | | | |
| | /cell_line="978SK" | | | | | | | |
| | /cell_type="fibroblast" | | | | | | | |
| | /note="selection using chloramphenicol; orientation of the sequence is from centromere to telomere" | | | | | | | |
| STS | 40651.41020 | | | | | | | |
| | /note="STS marker D8S1724" | | | | | | | |
| BASE COUNT | 52344 a 32042 c 31175 g 51271 t | | | | | | | |
| ORIGIN | | | | | | | | |
| Query Match. | 11.6%; Score 129.8; DB 11; Length 166832; | | | | | | | |
| Best Local Similarity | 66.2%; Pred. No. 5.4e-26; | | | | | | | |
| Matches 280; Conservative | 0; Mismatches 132; Indels 11; Gaps 6; | | | | | | | |
| QY | 518 | GAACACTGTGATCAACTAGCAGCAGCAGCAGTGTCT-CCGAGTCAGACTATCATGATTAC | 576 | | | | | |
| Db | 47221 | GAACACCGTGATTAATAGCCAAATGCCAGAGCTCTACATGAGTCAGACTATTTAATGCC | 47280 | | | | | |
| QY | 577 | TGAGTGAAGTGTGCTGTCTATTGTACTTAACCATGCCGATGTTAGTGAA-----ATT | 630 | | | | | |
| Db | 47281 | TGCTTTGCCCTCTGCTGTCAATTATGTAATATGCCACCACTTGCCCTTGATGTTGAGT | 47340 | | | | | |

| | | | |
|----|-------|--|-------|
| QY | 631 | GCGTCACTGGGCTCACAACCATCTTGGTATCCAGTTATTCCTACTGAATTGAGATTTCCTG | 690 |
| Dd | 47341 | CCACCACCTGTGCCCTGCACACTTTGGGATCCAATTATTCCCACCTGCATTTAAATTTGTA | 47400 |
| Qy | 691 | CTTCAGTGTACGCCAATCCCCAC-ATAATTTCTGACCTACAGAGGTAGGAGTACATATAGC | 749 |
| Dd | 47401 | GTTGAGTAGCTGTGGTTCTCTGCTGTTAGATCTGGCATACAGAGAAGAACATCACAGAGC | 47460 |
| QY | 750 | TCTTCAAAGATGCTGTTACTCCTCCCTCACAATAATTCATTCTC-CGTGTAGTGAAGGTH | 808 |
| Dd | 47461 | TCTTCAAAGAGGAGGTGCTGCCCTCACAAATCTATTTCACAAAAGTGTGTAAGGHTA | 47520 |
| QY | 809 | C GCCCTCTGAGGCTCCCCAGG GTGGGTGTCAGG TC - ACAATG ATGA TGTATGAT CTHG | 867 |
| Dd | 47521 | TGCTCTCTGGACCCTCCACAGCTGGGATGATAGGTCTTAAGTGCATTAATCCACTCTIAGCA | 47580 |
| QY | 868 | TTCCCATTACCCAAAGCCTTT-AAATCCCTCATGCTCAGTACAGCAGGGCAGGCTIAGCA | 926 |
| Dd | 47581 | TCCCAATCTCCCTAAGCCTTTGGATTTCTTCCTCTACATTAACCAAGGAGATCAGSCA | 47640 |
| QY | 927 | TTTT 929 | |
| Dd | 47641 | TTTT 47643 | |

| | |
|------------|--|
| RESULT | 13 |
| AC005906/c | |
| LOCUS | |
| DEFINITION | AC005906 185952 bp DNA PRI 30-JAN-1999 Homo sapiens 12p13.3 BAC RPlc11-429A20 (Roswell Park Cancer Institute Human BAC Library) complete sequence. |
| ACCESSION | AC005906 |
| NID | 94165009 |
| VERSION | AC005906.1 GI:4165009 |
| KEYWORDS | HTG. |
| SOURCE | human. |
| ORGANISM | Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 185952) |
| REFERENCE | Muzny,D., Atenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H., Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S., Kondejewski,N., Lau,S., Leal,B., Lee,E., Licharge,O., Liu,W., Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A., Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W., zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A. Direct Submission |
| TITLE | Unpublished |
| JOURNAL | 2 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (03-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 3 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (30-DEC-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 4 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (20-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 5 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | Baylor Plaza, Houston, TX 77030, USA |

COMMENT

On Jan 20, 1999 this sequence version replaced g1:4075596.
 INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smilt and P. Green published.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Summary Statistics

| | |
|--|------------|
| Contig length: | 185961 |
| Phrap values in estimate: | 98505 |
| Average error rate (BCM-Phrap estimate): | 5.51369e-0 |
| Fraction of Phrap values less than 40 : | 0.00108624 |
| Number of consensus changing edits: | 4 |
| Number of N's in consensus : | 0 |

| Position | Original+Context | Consensus changing edits | Edited+Context |
|----------|-----------------------|--------------------------|-----------------------|
| 9492 | aaagtaatc(t)tttttttt | | aaagtaatc(c)tttttttt |
| 36770 | tgggtgtggg(n)agcgggag | | tgggtgtggg(g)agcgggag |
| 107439 | cttgcccttg(n)cgttaact | | cttgcccttg(a)cgttaact |
| 109251 | gtcccgagc(n)ccctactac | | gtcccgagc(c)ccctactac |

```
----- Bases with BCM-Phrap value < 20-----
Quality      Position      Surrounding Sequence
```

----- Distribution of Quality < 40 Bases

| # | bases |
|-----|-------|
| 100 | |
| 90 | |
| 80 | |
| 70 | |
| 60 | |
| 50 | |
| 40 | |
| 30 | |
| 20 | |

|||||
Db 361 GATGCTGTGATTTGCCATCCAGTCCGAGACTGTGGAGGCTGGGAGTGTGAGAAGCTTTCC 420
QY 421 CAACCCCTGGCAGGGTGTACCATTTCCGCAACTTCCAGTGCAGAGGAGCTCCTGCTGCATC 480
Db 421 CAACCCCTGGCAGGGTGTACCATTTCCGCAACTTCCAGTGCAGAGGAGCTCCTGCTGCATC 480
QY 481 CTCACCTGGTGTCTACTACTGTCTACTGTGCATCACCAGCACTGTGATCAACTAGCCAG 540
Db 481 CTCACCTGGTGTCTACTACTGTCTACTGTGCATCACCAGCACTGTGATCAACTAGCCAG 540
QY 541 CACCATAGTCTCCGAGTCAAGTCAATCATGATTAATGATGTTGACTGTGCTGTCTATGTT 600
Db 541 CACCATAGTCTCCGAGTCAAGTCAATCATGATTAATGATGTTGACTGTGCTGTCTATGTT 600
QY 601 ACTAACCATGCCGATGTTAGTGAATTAAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
Db 601 ACTAACCATGCCGATGTTAGTGAATTAAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
QY 661 CAGTTATCTCTACTGTAATGAGATTCTCTGCTCAGTGTGAGCCATTTCCACATTAATTTC 720
Db 661 CAGTTATCTCTACTGTAATGAGATTCTCTGCTCAGTGTGAGCCATTTCCACATTAATTTC 720
QY 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCACAATA 780
Db 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCACAATA 780
QY 781 TTCATTTCTCTCTGTGTAGTGAAGGTGCGCCCTCTGAGCCTCCAGGGTGGTGTGCA 840
Db 781 TTCATTTCTCTCTGTGTAGTGAAGGTGCGCCCTCTGAGCCTCCAGGGTGGTGTGCA 840
QY 841 GGTCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 841 GGTCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
QY 901 CTCAGTACACGAGGAGGAGTCTAGCATTTCTTCATTTAGTGTATGCTGTCCATTCATGCA 960
Db 901 CTCAGTACACGAGGAGGAGTCTAGCATTTCTTCATTTAGTGTATGCTGTCCATTCATGCA 960
QY 961 ACCACCTCAGAGCTCTGGATCTCTGCTCTAGTGTGAGCTCCTGCATGCTGCTCTGGG 1020
Db 961 ACCACCTCAGAGCTCTGGATCTCTGCTCTAGTGTGAGCTCCTGCATGCTGCTCTGGG 1020
QY 1021 GAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 1021 GAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 1081 TTATTAACAAGAGCTGTGATGTTAAAAA 1119
Db 1081 TTATTAACAAGAGCTGTGATGTTAAAAA 1119

RESULT 2

V61253
ID V61253 standard; cDNA; 1119 BP.
AC V61253;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN MO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
claim 3; Page 108-109; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 100.0%; Score 1119; DB 1; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCAGTCCGAGCCCTGGCAGCGCGCACTGTGATGGAAGCAATGTTCTGCTCGGGC 60
Db 1 GCGCAGTCCGAGCCCTGGCAGCGCGCACTGTGATGGAAGCAATGTTCTGCTCGGGC 60
QY 61 GTCCTGTGTCATCCGAGTGGTGTCTGTCAGCCGCACTGTTCCAGAACTCTACACC 120
Db 61 GTCCTGTGTCATCCGAGTGGTGTCTGTCAGCCGCACTGTTCCAGAACTCTACACC 120
QY 121 ATCGGGCTGGGCTGCACAGTCTTGAGCGCCAGCAAGAGCCAGGAGCCAGATGTTGAG 180
Db 121 ATCGGGCTGGGCTGCACAGTCTTGAGCGCCAGCAAGAGCCAGGAGCCAGATGTTGAG 180
QY 181 GCCAGCTCTCCGTACGCGCAGCAGAGTACACAGACCCCTTGCTCGTAAAGACCTCATG 240
Db 181 GCCAGCTCTCCGTACGCGCAGCAGAGTACACAGACCCCTTGCTCGTAAAGACCTCATG 240
QY 241 CTCATCAAGTTGAGCAATCCGTTCCGAGTCTGACACCATCCGGAGCATGCAATTGCT 300
Db 241 CTCATCAAGTTGAGCAATCCGTTCCGAGTCTGACACCATCCGGAGCATGCAATTGCT 300
QY 301 TCGCAGTGGCCTACCGCGGGGAATCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 301 TCGCAGTGGCCTACCGCGGGGAATCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 361 GATGCTGTGATGTCATTCAGTCCAGAGTGTGGAGGCTGGGAGTGTGAAGACTTTCC 420
Db 361 GATGCTGTGATGTCATTCAGTCCAGAGTGTGGAGGCTGGGAGTGTGAAGACTTTCC 420
QY 421 CAACCTGGCAGGGTGTACATTTCCGCAACTTCCAGTGCAGAGAGCTGCTGCATC 480
Db 421 CAACCTGGCAGGGTGTACATTTCCGCAACTTCCAGTGCAGAGAGCTGCTGCATC 480
QY 481 CTCACCTGGTGTCTACTACTGTCTACTGTGCATCACCAGCACTGTGATCAACTAGCCAG 540
Db 481 CTCACCTGGTGTCTACTACTGTCTACTGTGCATCACCAGCACTGTGATCAACTAGCCAG 540
QY 541 CACCATAGTCTCCGAGTCAAGTCAATGATTAATGATGTTGACTGTGCTGTCTATGTT 600
Db 541 CACCATAGTCTCCGAGTCAAGTCAATGATTAATGATGTTGACTGTGCTGTCTATGTT 600
QY 601 ACTAACCATGCCGATGTTAGTGAATTAAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
Db 601 ACTAACCATGCCGATGTTAGTGAATTAAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
QY 661 CAGTTATCTCTACTGTAATGAGATTCTCTGCTCAGTGTGAGCCATTTCCACATTAATTTC 720
Db 661 CAGTTATCTCTACTGTAATGAGATTCTCTGCTCAGTGTGAGCCATTTCCACATTAATTTC 720
QY 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCACAATA 780
Db 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCACAATA 780
QY 781 TTCATTTCTCTCTGTGTAGTGAAGGTGCGCCCTCTGAGAGCTCCAGGGTGGTGTGCA 840
Db 781 TTCATTTCTCTCTGTGTAGTGAAGGTGCGCCCTCTGAGAGCTCCAGGGTGGTGTGCA 840
QY 841 GGTCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 841 GGTCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900

OY 901 CTCAGTACACCAGGCGAGGTCTAGCATTTCTTATTAGTATAGCTGTCCATTCATGCA 960
|
DB 901 CTCAGTACACCAGGCGAGGTCTAGCATTTCTTATTAGTATAGCTGTCCATTCATGCA 960
OY 961 ACCACTCAGGACTCCTGATTTCTCTGCTAGTGTAGCTCCTGCATGCTGCCCTTGGG 1020
|
DB 961 ACCACTCAGGACTCCTGATTTCTCTGCTAGTGTAGCTCCTGCATGCTGCCCTTGGG 1020
OY 1021 GAGGTGAGGAGAGGGGCCCATGTTCAATGGGATCTGTGAGTGTTAACACATTAAGTGC 1080
|
DB 1021 GAGGTGAGGAGAGGGGCCCATGTTCAATGGGATCTGTGAGTGTTAACACATTAAGTGC 1080
OY 1081 TTAATTAACAGAGCTGTGATGTTAAAAA 1119
|
DB 1081 TTAATTAACAGAGCTGTGATGTTAAAAA 1119

RESULT 3
V58645
ID V58645 standard; cDNA; 1265 BP.
AC V58645;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.3%; Score 361.4; DB 1; Length 1265;
Best Local Similarity 99.7%; Pred. No. 6.5e-98;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CGCACTCGACGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTGGGCG 61
|
DB 8 CGCACTCGACGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTGGGCG 67
OY 62 TCCTGTGCATCCGCAAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCA 121
|
DB 68 TCCTGTGCATCCGCAAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCA 127
OY 122 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAAG 181
|
DB 128 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAAG 187
OY 182 CCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTGCTAAGCACTCATGC 241
|
DB 188 CCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTGCTAAGCACTCATGC 247
OY 242 TCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCAATCCGAGCATCAGATTGCTT 301
|
DB 248 TCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCAATCCGAGCATCAGATTGCTT 307

OY 302 CGCAGTGCCTTACCGCGGGGAACCTTGCCCTGCTTCTGCTGGGGTCTGTCGGAACG 361
|
DB 308 CGCAGTGCCTTACCGCGGGGAACCTTGCCCTGCTTCTGCTGGGGTCTGTCGGAACG 367
OY 362 ATG 364
|
DB 368 GTG 370

RESULT 4
V61250
ID V61250 standard; cDNA; 1265 BP.
AC V61250;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE2.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 105-106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.3%; Score 361.4; DB 1; Length 1265;
Best Local Similarity 99.7%; Pred. No. 6.5e-98;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CGCACTCGACGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTGGGCG 61
|
DB 8 CGCACTCGACGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTGGGCG 67
OY 62 TCCTGTGCATCCGCAAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCA 121
|
DB 68 TCCTGTGCATCCGCAAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCCTACACCA 127
OY 122 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAAG 181
|
DB 128 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAAG 187
OY 182 CCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTGCTAAGCACTCATGC 241
|
DB 188 CCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTGCTAAGCACTCATGC 247
OY 242 TCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCAATCCGAGCATCAGATTGCTT 301
|
DB 248 TCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACCAATCCGAGCATCAGATTGCTT 307
OY 302 CGCAGTGCCTTACCGCGGGGAACCTTGCCCTGCTTCTGCTGGGGTCTGTCGGAACG 361
|
DB 308 CGCAGTGCCTTACCGCGGGGAACCTTGCCCTGCTTCTGCTGGGGTCTGTCGGAACG 367
OY 362 ATG 364
|
DB 368 GTG 370

RESULT 5


```
V58647
ID V58647 standard; cDNA; 1167 BP.
AC V58647;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KM therapy; ss.
OS Homo sapiens.
FH Key
FT CDS Location/Qualifiers
    28..645
        /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W693388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;
```

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 31.5%; | Score 352.2; | DB 1; | Length 1167; |
| Best Local Similarity | 96.3%; | Pred. No. 3.4e-95; | | |
| Matches 360; | Conservative 0; | Mismatches 14; | Indels 0; | Gaps 0; |

[illegible]

| RESULT | 6 |
|--------|---------------------------------|
| V61252 | |
| ID | V61252 standard; cDNA; 1167 BP. |
| AC | W61252;... |

DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
P-PSDB; W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
of a prostate tumour protein. The encoded immunogen, or the DNA itself
can be used as a vaccine for the treatment of prostate cancer. The DNA
was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 31.5% | Score 352.2; | DB 1; | Length 1167; |
| Best Local Similarity | 96.3%; | Pred. No. 3.4e-95; | | |
| Matches 360; Conservative | 0; | Mismatches 14; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 8 | CGCAGCCCTGGCAGCGGCACTGTGTCATGSAAAACGAATTGTTCTGCTCGGGCGTCTTG | 67 |
| | | | |
| Dd | 2 | CGCAGCCCTGGCAGCGGCACTGTGTCATGSAAAACGAATTGTTCTGCTCGGGCGTCTTG | 61 |
| QY | 68 | TGCATCCGCAGTGGGTGCTGTCAGCCCGCACACTGTTCCAGAActCTTACACCATCGGC | 127 |
| | | | |
| Dd | 62 | TGCATCCGCAGTGGGTGCTGTCAGCCCGCACACTGTTCCAGAActCTTACACCATCGGC | 121 |
| QY | 128 | TGGGCGTCACAGTCTTGAGGCCGACCAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 187 |
| | | | |
| Dd | 122 | TGGGCGTCACAGTCTTGAGGCCGACCAGAGCCAGGAGCCAGATGTTGAGGCCAGCC | 181 |
| QY | 188 | TCTCCGTACGGCACCCAGAGTACAACACAGACCCTTGCTCGCTAACGACCCTCATGTCTATCA | 247 |
| | | | |
| Dd | 182 | TCTCCGTACGGCACCCAGAGTACAACACAGACTCTTGCTCGCTAACGACCCTCATGTCTATCA | 241 |
| QY | 248 | AGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGCATTTGCTTCGCAGT | 307 |
| | | | |
| Dd | 242 | AGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGCATTTGCTTCGCAGT | 301 |
| QY | 308 | GCCCTACCGCGGGGAActTCTTGCCCTGTTCTGGCTGGGCTCTGCTGGCAAGATGCTG | 367 |
| | | | |
| Dd | 302 | GCCCTACCGCGGGGAActTCTTGCCCTGTTCTGGCTGGGCTCTGCTGGCAAGCGCAGAA | 361 |
| QY | 368 | TGATTGCCATCCAG | 381 |
| | | | |
| Dd | 362 | TGCTTACCGTGTG | 375 |

| | |
|--------|--|
| RESULT | 7 |
| V37495 | |
| ID | V37495 standard; DNA; 871 BP. |
| AC | V37495; |
| DT | 07-SEP-1998 (first entry) |
| DE | Human prostate-specific kallikrein (HPSK) encoding DNA. |
| KW | prostate-specific kallikrein; HPSK; prostate carcinoma; human; |
| OS | benign prostate hyperplasia; diagnosis; drug screening; PSK; ss. |
| FH | Homo sapiens. |
| FT | Key |
| FT | Location/Qualifiers |
| FT | 31..777 |
| FT | /*tag= a |
| FT | /transl_except= (pos:367..369, aa:Xaa) |
| FT | /transl_except= (pos:412..414, aa:Xaa) |

FT /transl_except- (pos:424. .426, aa:Xaa)
FT /product= "HPSK protein"
FT /note= "Xaa = unknown"
PN WO9820117-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goli SK;
DR WPI: 98-286933/25.
DR P-PSDB; W60592.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 5; Fig 1A-C; 68BP; English.
CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

Query Match 31.3%; Score 350.4; DB 1; Length 871;
Best Local Similarity 95.0%; Pred. No. 1e-94;
Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GCACTCGAGCCCTGGCGAGCGGCGACTGTGCAATGAAACGAATGTTCTGCTCGGCGT 62
DB 129 GCATTCCAGCCCTGGCGAGCGGCGACTGTGCAATGAAACGAATGTTCTGCTCGGCGT 188
QY 63 CCTGTGCATCCGAGTGGGTGCTGTACGCGCACACTGTTCCAGAACTCCTACACCAT 122
DB 189 CCTGTGCATCCGAGTGGGTGCTGTACGCGCACACTGTTCCAGAACTCCTACACCAT 248
QY 123 CGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGC 182
DB 249 CGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGC 308
QY 183 CAGCCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATGCT 242
DB 309 CAGCCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATGNT 368
QY 243 CATCAAGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATGCTTC 302
DB 369 CATCAAGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATGNTTC 428
QY 303 GCAGTGCCTACCGGGGGAACCTTGCTGCTGTTCTGGCTGGGGTCTGCTGGCGAAGA 362
DB 429 GCAGTGCCTACCGGGGGAACCTTGCTGCTGTTCTGGCTGGGGTCTGCTGGCGAAGC 488
QY 363 TGCTGTGATGCCATCCAG 381
DB 489 CAGAATGCCCTACCGTGTG 507

RESULT 8
V11855
ID V11855 standard; cDNA; 1386 BP.
AC V11855;
DT 11-SEP-1998 (first entry)
DE Homo sapiens Tub Interactor (hTI-1) gene.
KW serine protease; tub interactor; treatment; obesity; cachexia;
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW neurodegenerative disease; Alzheimer's disease; drug screening;
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
OS Homo sapiens.

FH Key Location/Qualifiers
FT CDS 2..701
FT /*tag= a
FT /product= hTI-1 protein
FT /note= "putative serine protease"
PN WO9812302-A1.
PD 26-MAR-1998.
PF 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errada PR, Gimeno CJ;
DR WPI: 98-217246/19.
DR P-PSDB; W59129.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 10; Fig 1; 120BP; English.
CC The sequence is that of the Tub Interactor gene hTI-1 which
CC codes for a putative serine protease. TI genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 30.4%; Score 339.8; DB 1; Length 1386;
Best Local Similarity 93.9%; Pred. No. 1.8e-91;
Matches 369; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 1 GCGCACTCGAGCCCTGGCGAGCGGCGACTGTGCAATGAAACGAATGTTCTGCTCGGCG 60
DB 14 GCGCACTCGAGCCCTGGCGAGCGGCGACTGTGCAATGAAACGAATGTTCTGCTCGGCG 73
QY 61 GTCCTGTGCATCCGAGTGGGTGCTGTACGCGCACACTGTTCCAGAA----- 111
DB 74 GTCCTGTGCATCCGAGTGGGTGCTGTACGCGCACACTGTTCCAGAAAGTGAAGTGCAG 133
QY 111 --CTCTACACCATCGGGCTGGGCTTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 168
DB 134 AGCTCTACACCATCGGGCTGGGCTTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 193
QY 169 CAGATGTGAGGCGACGCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCT 228
DB 194 CAGATGTGAGGCGACGCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCT 253
QY 229 AAGCACTCATGCTCATCAAGTTGACGAATCCGTGTCGAGTCTGACACCATCCGAGC 288
DB 254 AAGCACTCATGCTCATCAAGTTGACGAATCCGTGTCGAGTCTGACACCATCCGAGC 313
QY 289 ATCAGCATTCCTTCGAGTGCCTACCGGGGGAACCTTGCTGCTGTTCTTGCTGGGCT 348
DB 314 ATCAGCATTCCTTCGAGTGCCTACCGGGGGAACCTTGCTGCTGTTCTTGCTGGGCT 373
QY 349 CTGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
DB 374 CTGCTGGCGAAGCATGCTGTGATGCCATCCAG 406

RESULT 9
V58644
ID V58644 standard; cDNA; 1248 BP.
AC V58644;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DEL.

KW prostate tumour specific gene; human; prostate cancer; detection;
KM therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 217..696
PI /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69387.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 112; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
Best Local Similarity 93.9%; Pred. No. 3.4e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAAGCAATTGTTCTGCTCGGCG 61
Db 35 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAAGCAATTGTTCTGCTCGGCG 94
QY 62 TCCTGGTGATCCGCAAGTGGGTGCTGCAGCCGACACTGTTCCAGAA----- 111
Db 95 TCCTGGTGATCCGCAAGTGGGTGCTGCAGCCGACACTGTTCCAGAAAGTAGTGAGCA 154
QY 111 -CTCTACACCATCGGCGCTGGCGCTGCAGAGTCTTGAGGCCGACCAAGAGCCAGGAGCC 169
Db 155 GCTCTACACCATCGGCGCTGGCGCTGCAGAGTCTTGAGGCCGACCAAGAGCCAGGAGCC 214
QY 170 AGATGGTGAGGCCAGCCTCTCCGTACGCGCACCCAGAGTACACAGACCCTTGCTCGCTA 229
Db 215 AGATGGTGAGGCCAGCCTCTCCGTACGCGCACCCAGAGTACACAGACCCTTGCTCGCTA 274
QY 230 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 289
Db 275 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 334
QY 290 TCAGCATTCCTTCGCAAGTGGCCCTACCCGGGGGAACCTTGCTGCTTCTGGCTGGGCTC 349
Db 335 TCAGCATTCCTTCGCAAGTGGCCCTACCCGGGGGAACCTTGCTGCTTCTGGCTGGGCTC 394
QY 350 TGCTGGCGAAGCATGCTGTGATTGCCATCCAG 381
Db 395 TGCTGGCGAAGCATGCTGTGATTGCCATCCAG 426

RESULT 10
V61249
ID V61249 standard; cDNA; 1248 BP.
AC V61249;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DEL.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.

PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
Best Local Similarity 93.9%; Pred. No. 3.4e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAAGCAATTGTTCTGCTCGGCG 61
Db 35 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAAGCAATTGTTCTGCTCGGCG 94
QY 62 TCCTGGTGATCCGCAAGTGGGTGCTGCAGCCGACACTGTTCCAGAA----- 111
Db 95 TCCTGGTGATCCGCAAGTGGGTGCTGCAGCCGACACTGTTCCAGAAAGTAGTGAGCA 154
QY 111 -CTCTACACCATCGGCGCTGGCGCTGCAGAGTCTTGAGGCCGACCAAGAGCCAGGAGCC 169
Db 155 GCTCTACACCATCGGCGCTGGCGCTGCAGAGTCTTGAGGCCGACCAAGAGCCAGGAGCC 214
QY 170 AGATGGTGAGGCCAGCCTCTCCGTACGCGCACCCAGAGTACACAGACCCTTGCTCGCTA 229
Db 215 AGATGGTGAGGCCAGCCTCTCCGTACGCGCACCCAGAGTACACAGACCCTTGCTCGCTA 274
QY 230 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 289
Db 275 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 334
QY 290 TCAGCATTCCTTCGCAAGTGGCCCTACCCGGGGGAACCTTGCTGCTTCTGGCTGGGCTC 349
Db 335 TCAGCATTCCTTCGCAAGTGGCCCTACCCGGGGGAACCTTGCTGCTTCTGGCTGGGCTC 394
QY 350 TGCTGGCGAAGCATGCTGTGATTGCCATCCAG 381
Db 395 TGCTGGCGAAGCATGCTGTGATTGCCATCCAG 426

RESULT 11
X41114
ID X41114 standard; cDNA; 402 BP.
AC X41114;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:58.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IBI222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 99-153778/13.
DR P-PSDB: Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824bp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, hematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding of
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 30.2%; Score 338; DB 1; Length 402;
Best Local Similarity 96.8%; Pred. No. 3.4e-91;
Matches 360; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

OY 2 CGCACTCGAGCCCTGGAGGCGGCGACTGTCATGGAAGAAGATTGTTCTCGGGCG 61
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
23 CGCACTCGAGCCCTGGAGGCGGCGACTGTCATGGAAGAAGATTGTTCTCGGGCG 82
OY 62 TCCTGTGTCATCCGAGTGGGTGCTGTACGCCGACACTGTTCCAGAA----- 111
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
83 TCCTGTGTCATCCGAGTGGGTGCTGTACGCCGACACTGTTCCAGAAAGTGAATKAGA 142
OY 111 -CTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGAGCC 169
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
143 GCTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGAGCC 202
OY 170 AGATGGTGAGGCCAGCTCTCCGTACGGCACCAGAGTACACAGAGACCCTTGCTGCTA 229
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
203 AGATGGTGAGGCCAGCTCTCCGTACGGCACCAGAGTACACAGAGACCCTTGCTGCTA 262
OY 230 ACGACCTCATGCTCATCAAGTTGAGAGCAATCCGTGTCCAGTCTGACACCATCCGAGCA 289
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
263 ACGACCTCATGCTCATCAAGTTGAGAGCAATCCGTGTCCAGTCTGACACCATCCGAGCA 322
OY 290 TCAGCATTTGCTTCGAGTGCCTTACCGGGGGAAGTCTTGCTGCTGCTGGGTC 349
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
323 TCAGCATTTGCTTCGAGTGCCTTACCGGGGGAAGTCTTGCTGCTGCTGGGTC 382
OY 350 TGCTGGCGAAGC 361
Db ||||||||||||||||
383 TGCTGGCGAAGC 394

RESULT 12
ID V58646
AC V58646 standard; cDNA; 1459 BP.
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 09-FEB-1998; US-904809.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;
Best Local Similarity 97.8%; Pred. No. 6.3e-67;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 98 ACTGTTCCAGAACTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 157
Db ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
25 AGTAGTGACAGAGTCTCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 84
OY 158 AGCCAGGAGCCAGATGCTGAGGCCAGCCCTCTCCGTACGCCACCCAGAGTACACAGAC 217
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
85 AGCCAGGAGCCAGATGCTGAGGCCAGCCCTCTCCGTACGCCACCCAGAGTACACAGAC 144
OY 218 CCTGTCTGCTAACGACCTCATGCTCATCAAGTTGAGAGCAATCCGTGCTCGAGTGTACA 277
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
145 CCTGTCTGCTAACGACCTCATGCTCATCAAGTTGAGAGCAATCCGTGCTCGAGTGTACA 204
OY 278 CCATCCGAGCATGCAATGCTTGCAGTGCCCTACCGGGGAATCTTGCTGCTGTTT 337
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
205 CCATCCGAGCATGCAATGCTTGCAGTGCCCTACCGGGGAATCTTGCTGCTGTTT 264
OY 338 CTGGCTGGGCTGCTGCGAAGCATG 364
Db |||||||||||||||||||||||||
265 CTGGCTGGGCTGCTGCGAAGCATG 291

RESULT 13
ID V61251
AC V61251 standard; cDNA; 1459 BP.
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;
Best Local Similarity 97.8%; Pred. No. 6.3e-67;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 98 ACTGTTCCAGAACTCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 AGTGAGTGACAGAGCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAG 84
QY 158 AGCCAGGAGCCAGATGTTGAGGCCACGCTCTCCGTAAGCCACCCAGAGTACACAGAC 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 AGCCAGGAGCCAGATGTTGAGGCCACGCTCTCCGTAAGCCACCCAGAGTACACAGAC 144
QY 218 CCTTGTCTGCTAACGACCTCATGCTCATCAAGTTGAGGAATCCGTTCCGAGTCTGACA 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 CCTTGTCTGCTAACGACCTCATGCTCATCAAGTTGAGGAATCCGTTCCGAGTCTGACA 204
QY 278 CCATCCGGAGCATCAGCATGCTTGGCAGTGCCCTACCGGGGGAACCTCTGCTGTT 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 CCATCCGGAGCATCAGCATGCTTGGCAGTGCCCTACCGGGGGAACCTCTGCTGTT 264
QY 338 CTGGCTGGGCTGCTGCGGACGATG 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 CTGGCTGGGCTGCTGCGGACGATG 291
```

RESULT 14
V58522
ID V58522 standard; cDNA; 234 BP.
AC V58522;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone p20.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 56; 141p; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 13.7%; Score 153.8; DB 1; Length 234;
Best Local Similarity 93.1%; Pred. No. 1.8e-36;
Matches 161; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 209 ACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 60
QY 269 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGAGTCCCTACCGGGGGAACCTCTT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGAGTCCCTACCGGGGGAACCTCTT 120
QY 329 GCCTGTTTCTGCTGGGCTGCTGCTGCGGAACGATGCTGTGATGCGCATCCAG 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCCTGTTTCTGCTGGGCTGCTGCTGCGGAACGATGCTGTGATGCGCATCCGCTG 173
```

RESULT 15
V61168
ID V61168 standard; cDNA; 234 BP.
AC V61168;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone p20.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 53-54; 130p; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 13.7%; Score 153.8; DB 1; Length 234;
Best Local Similarity 93.1%; Pred. No. 1.8e-36;
Matches 161; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 209 ACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 60
QY 269 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGAGTCCCTACCGGGGGAACCTCTT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGAGTCCCTACCGGGGGAACCTCTT 120
QY 329 GCCTGTTTCTGCTGGGCTGCTGCTGCGGAACGATGCTGTGATGCGCATCCAG 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCCTGTTTCTGCTGGGCTGCTGCTGCGGAACGATGCTGTGATGCGCATCCGCTG 173
```

Search completed: September 28, 1999, 12:28:48
Job time: 5348 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:34:10 ; Search time 161.06 Seconds
(without alignments)
636.943 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119
Sequence: 1 GCGCACTCGCAGCCCTGCA.....ATGTTAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-------------------|-------------------|
| 1 | 350.4 | 31.3 | 871 | 3 | US-08-744-026-2 | Sequence 2, Appl |
| 2 | 81 | 7.2 | 986 | 4 | US-08-557-146-1 | Sequence 1, Appl |
| 3 | 81 | 7.2 | 1089 | 5 | PCT-US96-04294-1 | Sequence 1, Appl |
| 4 | 81 | 7.2 | 1089 | 5 | PCT-US96-04294-3 | Sequence 3, Appl |
| 5 | 76.4 | 6.8 | 832 | 5 | PCT-US95-06157-5 | Sequence 5, Appl |
| 6 | 72.8 | 6.5 | 992 | 2 | US-08-358-782D-13 | Sequence 13, Appl |
| 7 | 72.6 | 6.5 | 1462 | 2 | US-08-358-782D-14 | Sequence 14, Appl |
| 8 | 72 | 6.4 | 200 | 3 | US-08-931-981A-1 | Sequence 1, Appl |
| 9 | 71.6 | 6.4 | 766 | 5 | PCT-US95-06157-9 | Sequence 9, Appl |
| 10 | 68 | 6.1 | 1729 | 4 | US-08-844-024-1 | Sequence 1, Appl |
| 11 | 63.6 | 5.7 | 760 | 5 | PCT-US95-06157-7 | Sequence 7, Appl |
| 12 | 62.4 | 5.6 | 833 | 4 | US-08-790-137-2 | Sequence 2, Appl |
| 13 | 58.4 | 5.2 | 2259 | 4 | US-08-845-998-3 | Sequence 3, Appl |
| 14 | 56.8 | 5.1 | 2218 | 4 | US-08-845-998-5 | Sequence 5, Appl |
| 15 | 50.8 | 4.5 | 957 | 3 | US-08-684-862-11 | Sequence 11, Appl |
| 16 | 49.6 | 4.4 | 1454 | 4 | US-08-467-155A-2 | Sequence 2, Appl |
| 17 | 49.6 | 4.4 | 1454 | 4 | US-08-628-198-2 | Sequence 2, Appl |
| 18 | 49.6 | 4.4 | 1454 | 5 | PCT-US96-07343-2 | Sequence 12, Appl |
| 19 | 46 | 4.1 | 840 | 3 | US-08-684-862-12 | Sequence 2, Appl |
| 20 | 45.6 | 4.1 | 732 | 3 | US-08-361-395-2 | Sequence 42, Appl |
| 21 | 45 | 4.0 | 111 | 3 | US-08-454-720A-42 | Sequence 3, Appl |
| 22 | 40.8 | 3.6 | 925 | 1 | US-07-990-301A-3 | Sequence 14, Appl |
| 23 | 40.4 | 3.6 | 7218 | 2 | US-08-232-463-14 | Sequence 1, Appl |
| 24 | 40.2 | 3.6 | 734 | 3 | US-08-650-129-1 | Sequence 2, Appl |
| 25 | 40.2 | 3.6 | 821 | 3 | US-08-650-129-2 | Sequence 3, Appl |
| 26 | 40.2 | 3.6 | 866 | 3 | US-08-650-129-3 | Sequence 8, Appl |
| 27 | 38.4 | 3.4 | 1096 | 3 | US-08-684-862-8 | Sequence 1, Appl |
| 28 | 36.4 | 3.3 | 1554 | 3 | US-08-469-486-1 | Sequence 3, Appl |
| 29 | 36 | 3.2 | 970 | 1 | US-08-148-910-3 | Sequence 14, Appl |
| 30 | 36 | 3.2 | 2033 | 1 | US-08-148-910-14 | Sequence 3, Appl |
| 31 | 36 | 3.2 | 970 | 2 | US-08-448-937A-3 | Sequence 14, Appl |
| 32 | 36 | 3.2 | 2033 | 2 | US-08-448-937A-14 | Sequence 1, Appl |
| 33 | 35.8 | 3.2 | 959 | 4 | US-08-568-031-1 | Sequence 1, Appl |
| 34 | 35.8 | 3.2 | 959 | 4 | US-08-966-319-1 | Sequence 3, Appl |
| 35 | 35.4 | 3.2 | 2743 | 1 | US-08-396-479B-3 | Sequence 3, Appl |
| 36 | 35.4 | 3.2 | 2743 | 2 | US-08-818-823-3 | Sequence 1, Appl |
| 37 | 35.4 | 3.2 | 2749 | 4 | US-08-124-981A-1 | Sequence 1, Appl |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 38 | 35.4 | 3.2 | 2853 | 5 | PCT-US94-07297-36 | Sequence 36, Appl |
| 39 | 35 | 3.1 | 807 | 2 | US-08-270-584A-1 | Sequence 1, Appl |
| 40 | 35 | 3.1 | 807 | 4 | US-08-765-192-1 | Sequence 1, Appl |
| 41 | 34.8 | 3.1 | 867 | 1 | US-07-990-301A-1 | Sequence 9, Appl |
| 42 | 34.2 | 3.1 | 1333 | 3 | US-08-684-862-9 | Sequence 1, Appl |
| 43 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-1 | Sequence 3, Appl |
| 44 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-3 | Sequence 5, Appl |
| 45 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-5 | |

ALIGNMENTS

RESULT 1
US-08-744-026-2
Sequence 2, Application US/08744026
Patent No. 5786148
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
TITLE OF INVENTION: KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,026
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0154 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 871 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-744-026-2

Query Match 31.3%; Score 350.4; DB 3; Length 871;
Best Local Similarity 95.0%; Pred. No. 5.6e-103;
Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GCACCTGCGAGCCCTGGCAGGCGGCACTGTCATGGAAGAATGTTCTGCTGGCGGT 62
DB 129 GCATTCGAGCCCTGGCAGGCGGCACTGTCATGGAAGAATGTTCTGCTGGCGGT 188
QY 63 CCTGTGTCATCCGAGTGGTGTGTCAGCCGACACACTGTTCCAGAACTCTACACCAT 122
DB 189 CCTGTGTCATCCGAGTGGTGTGTCAGCCGACACACTGTTCCAGAACTCTACACCAT 248

OY 123 CGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGTTGAGGC 182
|||
Db 249 CGGGCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGTTGAGGC 308
OY 183 CAGCCTCTCCGTACGGCAGCCAGAGTACAACAGACCCCTGCTGCTAACGACCTCATGCT 242
|||
Db 309 CAGCCTCTCCGTACGGCAGCCAGAGTACAACAGACCCCTGCTGCTAACGACCTCATGCT 368
OY 243 CATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATGACATTGCTTC 302
|||
Db 369 CATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATGACATTGCTTC 428
OY 303 GCAGTGGCCCTACCGCGGGGGAACCTCTGCTCGTTCTGGCTGGGGCTCTGCTGGCGAAGCA 362
|||
Db 429 GCAGTGGCCCTACCGCGGGGGAACCTCTGCTCGTTCTGGCTGGGGCTCTGCTGGCGAAGCA 488
OY 363 TGCTGTGATGGCCATCCAG 381
|||
Db 489 CAGATGCTTACCGTGTCTG 507

RESULT 2
US-08-557-146-1
; Sequence 1, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 986 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..786
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 25..90
; FEATURE:
; NAME/KEY: mat_peptide

; LOCATION: 112..783
; US-08-557-146-1
Query Match 7.2%; Score 81; DB 4; Length 986;
Best Local Similarity 53.7%; Pred. No. 1.8e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;
OY 6 CTCGACGCCCTGGCAGCGGGCAGCTGTCATGAGAAACGAATTGTCTGCTGGGCGTCT 65
|||
Db 141 CTCGACGCCCTGGCAGCGGGCAGCTGTCATGAGAAACGAATTGTCTGCTGGGCGTCT 200
OY 66 GGTGATCCGCAAGTGGTGTCTGTACGCCGACACTGTTTCCAGAACTCTACACCATCGG 125
|||
Db 201 GGTGATGAGCGGTGGTGTCTGTACGCCGACACTGCAAGATGAATGATACACCGTGCA 260
OY 126 GCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCCAG 185
|||
Db 261 CTTGGGCAAGTATACGCTGGGCGACAGAGAG-----CTCAGAGATCAAGGCTCT 311
OY 186 CTTCTCCGTACGGCAGCCAGAGTACAACAGACCCCTTGCTGCTAACGACCTCATGCTCAT 245
|||
Db 312 GAAGTCATTTCCGCCACCCCGGCTACTCCACAGACAGACCCATGTTAATGACCTCATGCTCGT 371
OY 246 CAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATGACATGCTTCCGA 305
|||
Db 372 GAAGTCATATAGCCAGGCCAGCGCTGTCTATGCTGAAGAAAGTCAAGCTGCCCTCCCG 431
OY 306 GTGCCCTACCGCGGGGGAACCTTGTGCTGTTCTGCTGGGCTGTGCTGGCGAAGCATGC 365
|||
Db 432 CTGCGAACCCTTGAGACCACTGTACTGTCTCCGGCTGGGGGACACTACACGAGCCGAGA 491
OY 366 TGTGA 370
|||
Db 492 TGTGA 496

RESULT 3
PCT-US96-04294-1/c
; Sequence 1, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US96-04294-1

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTCGAGCCCTGGGACGGGCGACCTGGTCATGAAACGAATGTTCTGCTCGGGCGTCT 65
DB 857 CTCACACCATGGGAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTCGGAGGCGTCT 798
QY 66 GGTGATCCGAGTGGGTGCTGTGAGCCGACACTGTTTCCAGAACTCTACACCAATCGG 125
DB 797 GGTCAATGAGGCTGGGTGCTCAGTGGCCGCACTGCAAGATGATGATACACCGTGA 738
QY 126 GCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCAGATGTGAGGCCAG 185
DB 737 CTGGGCGAGTATAGCTGGGCGACAGAGAG-----CTCAGAGGATCAAGGCTC 687
QY 186 CCTCTCCGATGGGACACCCAGAGTACACAGACCCCTGCTCGCTAACGACCTCATGCTCAT 245
DB 686 GAAGTCATTCGCGCCACCCCGGCTACTCCACACAGACCCCATGTTAATGACCTCATGCTCT 627
QY 246 CAAGTTGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGCATGCTGCGCA 305
DB 626 GAAGCTCAATAGACGAGGCGAGGCTGTCAATCATGTGTGAAGAAAGTCAAGGCTGCCCTCCG 567
QY 306 GTGCCCTACCGGGGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365
DB 566 CTGCGAACCCTTGAACACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 4

PCT-US96-04294-3/c
; Sequence 3, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:

; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
PCT-US96-04294-3

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTCGAGCCCTGGGACGGGCGACCTGGTCATGAAACGAATGTTCTGCTCGGGCGTCT 65
DB 857 CTCACACCATGGGAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTCGGAGGCGTCT 798
QY 66 GGTGATCCGAGTGGGTGCTGTGAGCCGACCCGACACTGTTTCCAGAACTCTACACCAATCGG 125
DB 797 GGTCAATGAGGCTGGGTGCTCAGTGGCCGCACTGCGCAATGATGATGATGATGATGATGATGAT 738
QY 126 GCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGAGGCCAG 185
DB 737 CTGGGCGAGTATAGCTGGGCGACAGAGAG-----CTCAGAGGATCAAGGCTC 687
QY 246 CAAGTTGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGCATGCTGCGCA 305
DB 626 GAAGCTCAATAGACGAGGCGAGGCTGTCAATCATGTGTGAAGAAAGTCAAGGCTGCCCTCCG 567
QY 306 GTGCCCTACCGGGGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365
DB 566 CTGCGAACCCTTGAACACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 5

PCT-US95-06157-5
; Sequence 5, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25


```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06157
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Raasch, Kevin W.
: REGISTRATION NUMBER: 35,561
: REFERENCE/DOCKET NUMBER: 150.148W01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: TELEFAX: 612-339-3061
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 10..792
PCT-US95-06157-5

```

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 6.88; | Score 76.4; | DB 5; | Length 832; |
| Best Local Similarity | 63.08; | Pred. No. 4.8e-15; | | |
| Matches 136; Conservative | 0; | Mismatches 76; | Indels 4; | Gaps 1; |

| | | | |
|----|-----|--|-----|
| QY | 311 | CTACCGGGGAACTCTTGCCCTCGTTTCTGGCCTGGGGTCTGCTGGCGAAGCATGCTGTGA | 370 |
| | | | |
| Db | 5 | CCAGCATGTGGAACTGGTTCCTCTCCATGCCCTTGTCTGTGGGTHCACTGGTGCCGTGC | 64 |
| QY | 371 | TTCGCATCCAGTCCCAGACTGTGGAGCGTGGAGTGTGAGAAGCTTTCCCAACCCTGGC | 430 |
| | | | |
| Db | 65 | CCCTCATCCAGTCTCGGATTGTGGAGCGTGGAGTGTGAGAAGCATTTCCCAACCCTGGC | 124 |
| QY | 431 | AGG----GTTGTACCAATTTCCGCACTTCCAGTGC AAGSAGCTTCCTGCATCCTCACT | 486 |
| | | | |
| Db | 125 | AGGTGGCTGTGTACAGTCATGAGTAGGGCACACTGTGGGGGTGTCTGTTGACACCCCACT | 184 |
| QY | 487 | GGGTGCCTACTACTGCTCACTGCATCACCCGGAACA | 522 |
| | | | |
| Db | 185 | GGGTGCTCACAGCTGCCCAATTGCCCTAAGAGAATA | 220 |

RESULT 6
US-08-358-782D-13
; Sequence 13, Application US/08358782D
; Patent No. 5674682

```

; GENERAL INFORMATION:
;
; APPLICANT: Croce, Carlo
;
; APPLICANT: Gomella, Leonard
;
; APPLICANT: Mulholland, S. Grant
;
; APPLICANT: Moreno, Jose
;
; APPLICANT: Fischer, Rainer
;
; TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
;
; NUMBER OF SEQUENCES: 14
;
; CORRESPONDENCE ADDRESS:
;

```

ADDRESS: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
ZIP: 19103

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,782D
; FILING DATE: 15-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Beardell, Lori Y.
: REGISTRATION NUMBER: 34,293
: REFERENCE/DOCKET NUMBER: TJU-1327
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 992 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
:
US-08-358-782D-13

```

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------|
| Query Match | 6.5%; | Score 72.8; | DB 2; | Length 992; |
| Best Local Similarity | 59.9%; | Pred. No. 7.7e-14; | | |
| Matches 139; | Conservative | 0; | Mismatches 89; | Indels 4; |
| | | | | Gaps 1; |

| | | | |
|----|-----|---|-----|
| QY | 303 | GCA GTGCCCTACCGGGGGAAGACTCTTGCCTCGTTTCTGTGCTGGGGGTCTGCTGGCCAGCA | 362 |
| | | | |
| Db | 31 | GCTGTGTTNMCACCATGTGGGTCCCGGTGTCTTCCTCACCCCTGTCCGTGACGTGATTTGG | 90 |
| QY | 363 | TGCTGTGATTGCCATCCAGTCCCAGACTGTGGAGGCTGGGAGTGAGAAGCTTTCCCA | 422 |
| | | | |
| Db | 91 | TGCTGCACCCCTCATTCTGTCTCGAATTGTGGAGGCTGGGAGTGCGAGAAGCATTTCCA | 150 |
| QY | 423 | ACCCTGGCAGG----GTTGTACCATTTGGGCAACTTCCAGTGGCAAGACGTCTGTGCA | 478 |
| | | | |
| Db | 151 | ACCCTGGCAGGTCTTGTAGCCCTCTCGTGGCAGGGCAGTCTGCGGGGTGTTCTGTGTGCA | 210 |
| QY | 479 | TCCTCACTGGGTGCTCACTACTGCTCACTGCATCACC CGGAACACTGTGATC | 530 |
| | | | |
| Db | 211 | CCCCCAGTGGGTCTCTCAGACGTGCCCACTGCATCAGGAACA AAGCGTGATC | 262 |

RESULT 7
US-08-358-782D-14
; Sequence 14, Application US/08358782D
; Patent No. 5674682

APPLICANT: Croce, Carlo
APPLICANT: Gomella, Leonard
APPLICANT: Mulholland, S. Grant
APPLICANT: Moreno, Jose
APPLICANT: Fischer, Rainer
TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

;; ADDRESS: Woodcock Washburn Kurtz Mackiewicz &
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
;

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,782D
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: TJU-1327
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439

Db 9 CATCCAGTCTCGGATTGTGGAGGCTGGAGTGTGAGAACATTCACACCTGGCAGGT 68
QY 434 ---GTTGACCATTTGCGCACTTCAGTGCAAGAGAGCTCCTGTCATCCTCAGTGGT 490
Db 69 GGCTGTGTACAGTCATGGATGGGACACTGTGGGGGTGTCCTGTGTGCACCCCAAGTGGT 128

QY 491 GCTCACTACTGCTCACTGCATCACCAGGACA 522
Db 129 GCTCACAGCTGCCCATTCCTAAAGAGAATA 160

RESULT 10

US-08-844-024-1

; Sequence 1, Application US/08844024

; Patent No. 5840494

; GENERAL INFORMATION:

; APPLICANT: Katz, Aaron E., et al.

; TITLE OF INVENTION: A Method For Molecular Staging Of

; TITLE OF INVENTION: Prostate Cancer

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/844,024

; FILING DATE: 18-APR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/229,391

; FILING DATE: 15-APR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 43677/JPW/TEP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 977-9550

; TELEFAX: (212) 664-0525

; TELEX: 422523 COOP UI

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1729 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 378..1088

; US-08-844-024-1

Query Match 6.1%; Score 68; DB 4; Length 1729;
Best Local Similarity 66.1%; Pred. No. 3.7e-12;
Matches 115; Conservative 0; Mismatches 55; Indels 4; Gaps 1;

QY 361 GATGCTGATGTCATCCAGCTGTGGAGGCTGGAGTGTGAGAAGCTTCC 420
Db 351 GGTGCTGACCCCTCATCTCTCTGGAATTTGGAGGCTGGAGTGCAGAAAGCATTC 410
QY 421 CAACCTGCGCAGG---GTTGACCATTTGCGCACTTCAGTGCAAGAGCTCTGCTG 476
Db 411 CAACCTGCGCAGGTTGTTAGCCTCTCTGTTGGCAGGCGCAGTCTGCGGCTGTTG 470

QY 477 CATCTCACTGGGTGCTCACTACTGCTCATGATCAACCCGGAACACTGTGATC 530
Db 471 CACCCCAAGTGGTCTCTACAGACTACCACACTGCATCAGAAACAAAGCGTGATC 524

RESULT 11

PCT-US95-06157-7

; Sequence 7, Application PC/TUS9506157

; GENERAL INFORMATION:

; APPLICANT: Mayo Foundation for Medical Education

; APPLICANT: and Research

; APPLICANT: Hybritech Incorporated

; APPLICANT: Tindall, Donald J.

; APPLICANT: Young, Charles Y.F.

; APPLICANT: Saeedi, Mohammed S.

; TITLE OF INVENTION: Recombinant HK2 Polypeptide

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.

; STREET: 3500 IDS Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/06157

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Raasch, Kevin W.

; REGISTRATION NUMBER: 35,561

; REFERENCE/DOCKET NUMBER: 150.148W01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-339-0331

; TELEFAX: 612-339-3061

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 760 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 7..720

; PCT-US95-06157-7

Query Match 5.7%; Score 63.6; DB 5; Length 760;
Best Local Similarity 58.6%; Pred. No. 5.9e-11;
Matches 130; Conservative 0; Mismatches 89; Indels 3; Gaps 1;

QY 3 GCACCTGCGAGCCCTGGCAGGCGGCACACTGTCATGGAACGAATTGTTCTGCTGGCGGT 62
Db 36 GCATTCCCAACCCCTGGCAGGTGGCTGTGTACAGTCATGATGGGCACACTGTGGGGTGT 95
QY 63 CCTGTGTCATCCGAGTGGGTGCTGTGACGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGC 122
Db 96 CCTGTGTCACCCCGAGTGGGTGCTGTGACGCCGACCAAGAGCCAGATGGCCATTAAGAAAGATAGCCAGGT 155
QY 123 CGGGTGGCCCTGCACAGTTTGAAGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGC 182
Db 156 CTGGTGGGTGGCACAACCTGTTTGAAGCCTGAAGACACAG---GCCAGAGGGTCCCTGT 212
QY 183 CAGCCTCTCCGTACGGCACCAGAGTACACAGACCCCTTGT 224
Db 213 CAGCCACAGCTTCCACACACCCGCTCTACATATAGAGCTTCT 254

| | | | |
|----|-----|---|-----|
| QY | 283 | CGGAGCATCAGCAT-----TgCTTGCAGTgCCCCtACCGCGGGGAACTCTTgCCTCGTT | 336 |
| | | | |
| Db | 419 | CACACGGTCACGCTgCCCCCTgCCTCGAGACCTTCCCCCGGGGATgCCTgCTGGGTC | 478 |
| QY | 337 | TCTGgCTGGGgTCTgCTgGCGAAGCATgCTgTgATTgCCATCCAGTCCAGACTgTGGA | 396 |
| | | | |
| Db | 479 | ACTGgCTGGGCGGCACTgGACATATATgTgGCTgTTGGGgACAGCGGGAGGCGCGGCCA | 538 |
| QY | 397 | GGCTGGGA | 404 |
| | | | |
| Db | 539 | GGTGGGCA | 546 |

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

| | | | | |
|-----------------------|-----------------|--------------------|------------|--------------|
| Query Match | 5.1%; | Score 56.8; | DB 4; | Length 2218; |
| Best Local Similarity | 51.4%; | Pred. No. 1.7e-08; | | |
| Matches 189; | Conservative 0; | Mismatches 167; | Indels 12; | Gaps 2 |

[illegible]

QY 105 --CCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCA 162
| | | | | | | | | | | | | | |
Db 239 CCGGACATCAAGGATCTGGCCGCCCTCAGGGTGCAACTGGGGAGCAGCACCTTTACTAC 298
163 GGGAGCCAGATGTGTGAGGCCAGCCCTCTCCGTACGGCACCAGAGTACAACAGACCCTTG 222
| | | | | | | | | | | | | | |
Db 299 CAGGACCAGCTGTGTCCGGTCAGCAGAGATCATCTGTGCACCCACAGTTCTACATCATCCAG 358
QY 223 CTCGCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCCGAGTCTGACACCATC 282
| | | | | | | | | | | | | | |
Db 359 ACCGGGGCGGACATCGCCCTGCTGTGAGCTGGAGAGAGCCCCGTGAACATCTCCAGCCACATC 418
QY 283 CGGAGCATCAGCAT-----TGCTTCGAGTGCCTACCGGGGGAACTCTTGCTCGTT 336
| | | | | | | | | | | | | | |
Db 419 CACACGTCACGCTGCCCCCCTGCTCGAGACCTTCCCCCGGGGATGCGGTGCTGGTTC 478
QY 337 TCTGGCTGGGGTCTGCTGGCGAAGATGCTGTGATTGCCATCCAGTCCAGACTGTGGGA 396
| | | | | | | | | | | | | | |
Db 479 ACTGGCTGGGGCGAGCTGGACATAATGTGGGTGTTGGGGACAGCGGGAGGCCGGGCCA 538
QY 397 GGCTGGGA 404
| | | | |
Db 539 GGTGGGCA 546

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


This Page Blank (uspto)

OM of: US-09-030-606-177 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL-frame+n2p.model -DEV-xlp
-O-/cgn2_1/USPTO_spool/US09030606/runat_24091999_171618_29883/app_query.fasta.1
-DB-SwissProt_37-QFMT-fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS-human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-177
Query length: 1119
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|-------------------------|-----------|--------|---------|-----|--------------------------------|
| SwissProt_37:KLK1_MACFA | 275.00 | 435.48 | 4.2e-17 | 257 | Q07276 macaca fascicularis (|
| SwissProt_37:SCCE_HUMAN | 274.50 | 434.82 | 4.6e-17 | 253 | P49862 homo sapiens (human). |
| SwissProt_37:KLK1_HUMAN | 272.50 | 431.25 | 7.0e-17 | 262 | P06870 homo sapiens (human). |
| SwissProt_37:KLK1_PAPHA | 270.50 | 428.15 | 1.1e-16 | 258 | Q28773 papio hamadryas (hama |
| SwissProt_37:PROS_HUMAN | 270.00 | 427.23 | 1.2e-16 | 261 | P07288 homo sapiens (human). |
| SwissProt_37:KLK2_HUMAN | 266.50 | 421.56 | 2.4e-16 | 261 | P20151 homo sapiens (human). |
| SwissProt_37:PROS_MACMU | 262.00 | 414.27 | 6.2e-16 | 261 | P33619 macaca mulatta (thesu |
| SwissProt_37:KLK3_MOUSE | 260.00 | 411.10 | 9.4e-16 | 259 | P15948 mus musculus (mouse). |
| SwissProt_37:KLK3_MOUSE | 255.00 | 402.92 | 2.7e-15 | 261 | P00756 mus musculus (mouse). |
| SwissProt_37:KLK4_RAT | 253.50 | 401.12 | 3.6e-15 | 244 | P36375 rattus norvegicus (rat) |
| SwissProt_37:KLKA_MOUSE | 250.50 | 395.63 | 6.8e-15 | 261 | P15946 mus musculus (mouse). |
| SwissProt_37:TRV3_CHICK | 249.00 | 393.68 | 9.2e-15 | 248 | Q90629 gallus gallus (chicke |
| SwissProt_37:TRV1_CANFA | 248.00 | 392.13 | 1.1e-14 | 246 | P06871 canis familiaris (dog |
| SwissProt_37:KLK2_RAT | 246.50 | 389.22 | 1.6e-14 | 259 | P00759 rattus norvegicus (rat) |
| SwissProt_37:KLK7_RAT | 245.50 | 387.53 | 1.9e-14 | 261 | P36373 rattus norvegicus (rat) |
| SwissProt_37:TRV4_RAT | 245.00 | 387.27 | 2.1e-14 | 246 | P32821 rattus norvegicus (rat) |
| SwissProt_37:TRVB_RAT | 245.00 | 387.27 | 2.1e-14 | 246 | P32822 rattus norvegicus (rat) |
| SwissProt_37:TRV1_BOVIN | 244.00 | 385.76 | 2.6e-14 | 243 | P00760 bos taurus (bovine). |
| SwissProt_37:KLK1_RAT | 244.00 | 385.10 | 2.6e-14 | 261 | P00758 rattus norvegicus (rat) |
| SwissProt_37:KLK9_MOUSE | 244.00 | 385.10 | 2.6e-14 | 261 | P15949 mus musculus (mouse). |
| SwissProt_37:KLK4_MOUSE | 243.50 | 384.47 | 2.9e-14 | 256 | P00757 mus musculus (mouse). |
| SwissProt_37:KLK8_RAT | 243.50 | 384.29 | 2.9e-14 | 261 | P36374 rattus norvegicus (rat) |
| SwissProt_37:TRYP_SQUAC | 243.00 | 384.70 | 3.1e-14 | 229 | P00764 squalus acanthias (sp |
| SwissProt_37:TRV3_RAT | 243.00 | 383.99 | 3.2e-14 | 247 | P08426 rattus norvegicus (rat) |
| SwissProt_37:TRV4_HUMAN | 242.50 | 381.25 | 3.7e-14 | 304 | P35030 homo sapiens (human). |
| SwissProt_37:KLK2_MOUSE | 242.00 | 381.86 | 4.0e-14 | 261 | P36369 mus musculus (mouse). |
| SwissProt_37:TRV1_RAT | 241.00 | 380.79 | 4.8e-14 | 246 | P00762 rattus norvegicus (rat) |
| SwissProt_37:KLK3_MOUSE | 241.00 | 380.24 | 4.9e-14 | 261 | P36368 mus musculus (mouse). |
| SwissProt_37:KLKB_RAT | 240.50 | 379.50 | 5.4e-14 | 259 | P36376 rattus norvegicus (rat) |
| SwissProt_37:TRV3_HUMAN | 240.00 | 379.13 | 6.0e-14 | 247 | P15951 homo sapiens (human). |
| SwissProt_37:TRV4_RAT | 240.00 | 379.13 | 6.0e-14 | 247 | P12788 rattus norvegicus (rat) |
| SwissProt_37:TRV1_CHICK | 239.00 | 377.47 | 7.3e-14 | 248 | Q90627 gallus gallus (chicke |
| SwissProt_37:TRV2_CHICK | 239.00 | 377.47 | 7.3e-14 | 248 | Q90628 gallus gallus (chicke |
| SwissProt_37:KLK2_CAVPO | 237.50 | 375.39 | 1.0e-13 | 239 | P12323 cavia porcellus (guin |
| SwissProt_37:TRV2_RAT | 237.00 | 374.31 | 1.1e-13 | 246 | P00763 rattus norvegicus (rat) |
| SwissProt_37:TRYP_MOUSE | 237.00 | 374.31 | 1.1e-13 | 246 | P07146 mus musculus (mouse). |
| SwissProt_37:TRV1_HUMAN | 237.00 | 374.27 | 1.1e-13 | 247 | P07477 homo sapiens (human). |
| SwissProt_37:ESTA_CANFA | 237.00 | 373.79 | 1.1e-13 | 260 | P09582 canis familiaris (dog |
| SwissProt_37:TRV2_XENLA | 236.00 | 372.76 | 1.4e-13 | 244 | P70059 xenopus laevis (afrio |
| SwissProt_37:KLK5_MOUSE | 236.00 | 372.13 | 1.4e-13 | 261 | P15945 mus musculus (mouse). |
| SwissProt_37:TRYP_PIG | 235.00 | 371.65 | 1.7e-13 | 231 | P00761 sus scrofa (pig). tryps |
| SwissProt_37:KLK1_MOUSE | 234.00 | 368.89 | 2.1e-13 | 261 | P15947 mus musculus (mouse). |

SwissProt_37:KLK6_MOUSE + 234.00 368.89 2.1e-13 261 P00755 mus musculus (mous
SwissProt_37:TRV2_SALSA + 232.00 366.79 3.1e-13 231 P35032 salmo salar (atlan
SwissProt_37:KLK8_MOUSE + 231.00 364.03 3.9e-13 261 P07628 mus musculus (mous

seq_name: SwissProt_37:KLK1_MACFA

| seq_documentation_block: | seq_documentation_block: | seq_documentation_block: |
|--|--------------------------|---------------------------|
| ID KLK1_MACFA | STANDARD; | PRT; 257 AA. |
| AC Q07276; | | |
| DT 01-OCT-1994 (REL. 30, CREATED) | | |
| DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE) | | |
| DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE) | | |
| DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN) | | |
| DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN). | | |
| GN KLK1. | | |
| OS MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY). | | |
| OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; | | |
| OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA. | | |
| RN [1] | | |
| RP SEQUENCE FROM N.A. | | |
| RX MEDLINE; 93305727. | | |
| RA LIN F.K., LIN C.H., CHOU C., CHEN K., LU H.S., BACHELLER B., | | |
| RA HERRERA C., JONES T., CHAO J., CHAO L.; | | |
| RT "Molecular cloning and sequence analysis of the monkey and human | | |
| tissue kallikrein genes." | | |
| RL BIOCHIM. BIOPHYS. ACTA 1173:325-328(1993). | | |
| CC -I- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS | | |
| CC IN KININOGEN TO RELEASE LYS-BRADYKININ. | | |
| CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-L-XAA BONDS IN | | |
| CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE | | |
| KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF | | |
| MET-L-XAA OR LEU-L-XAA. | | |
| CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | |
| TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY. | | |
| CC ----- | | |
| CC This SWISS-PROT entry is copyright. It is produced through a collaboration | | |
| between the Swiss Institute of Bioinformatics and the EMBL outstation - | | |
| the European Bioinformatics Institute. There are no restrictions on its | | |
| use by non-profit institutions as long as its content is in no way | | |
| modified and this statement is not removed. Usage by and for commercial | | |
| entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | |
| or send an email to license@isb-sib.ch). | | |
| CC ----- | | |
| DR EMBL; L10039; G293141; -. | | |
| DR PIR; S33772; S33772. | | |
| DR PROSITE; PS00134; TRYPSIN_HIS. 1. | | |
| DR PROSITE; PS00135; TRYPSIN_SER. 1. | | |
| DR PFAM; PF00089; trypsin; 1. | | |
| DR HSSP; P00752; IHIA. | | |
| KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS; | | |
| KW ZYMOGEN; SIGNAL. | | |
| FT SIGNAL | 1 | 18 |
| FT PROPEP | 19 | 24 |
| FT CHAIN | 25 | 257 |
| FT ACT_SITE | 62 | 62 |
| FT ACT_SITE | 116 | 116 |
| FT ACT_SITE | 209 | 209 |
| FT CARBOHYD | 90 | 90 |
| FT CARBOHYD | 99 | 99 |
| FT CARBOHYD | 101 | 101 |
| FT CARBOHYD | 105 | 105 |
| FT CARBOHYD | 160 | 160 |
| FT CARBOHYD | 162 | 162 |
| FT DISULFID | 31 | 169 |
| FT DISULFID | 47 | 63 |
| FT DISULFID | 148 | 215 |
| FT DISULFID | 180 | 194 |
| FT DISULFID | 205 | 230 |
| SEQUENCE | 257 AA; | 28237 MW; 82A709DC CRC32; |

alignment_scores: Quality: 275.00 Length: 158

Ratio: 2.865 Gaps: 3
Percent Similarity: 60.759 Percent Identity: 37.342

alignment_block:

US-09-030-606-177 x KLK1_MACFA ..

Align seg 1/1 to: KLK1_MACFA from: 1 to: 257

```

7 TCGAGCCCTGGCAGCGCGGCTGTCATGGAAGAATGTCTGCTC 56
|||||
32 SerGlnProTprGlnAlaLeuTyrHisPheSerThrPheGlnCysG1 48
57 GGGGCTCCTGTGTCATCCGAGTGGGTGCTGTACGCCACACTGTTCC 106
|||||
48 YGlyLeuValHisProGlnTprValLeuThrAlaAlaHisCysIles 65
107 AGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGGACCAA 156
|||||
65 erAspAsnTyrGlnLeuTprLeuGlyArgHisAsnLeu..PheAspAsp 80
157 GAGCCAGGAGCAGATGTGGAGCGCCAGCCTCTCCGTACGGCACCAGCA 206
|||||
81 GluAspThrAlaGlnPheValHisValSerGlnSerPheProHisProG1 97
207 GTACAACAGACCCCTGCTGCTGCTAAC..... 231
|||||
97 YPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspAspTyrS 114
232 .....GACSTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGAC 276
|||||
114 erHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGlnLeuThrAsp 130
277 ACCATCCGAGCATCAGCATGCTTGCAGTGCCTTACCGCGGGAATC 326
|||||
131 AlaValGlnValValGlnLeuProThrGlnGlnProGlnValGlySerTh 147
327 TTGCCTCGTTTCTGGCTGGGCT..... 348
|||||
147 rCysLeuAlaSerGlyTprGlySerIleGluProGlnAsnPheSerPheP 164
349 .....CTGCTGGCGGAACGATGCTGTG 369
|||||
164 roAspAspLeuGlnCysValAspLeuGlnLeuLeuProAsnAspGlnCys 180
370 ATTGCCATCCAGTCCCACTGTG 393
|||||
181 AlaLysAlaHisThrGlnLysVal 188

```

seq_name: SwissProt_37:SCCE_HUMAN

seq_documentation_block:

```

ID SCCE_HUMAN STANDARD; PRT; 253 AA.
AC P49862;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).
GN PRSS6 OR SCCE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=SKIN;
RX MEDLINE; 94308225.
RA HANSSON L., STROEMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A.,
RA EGELRUD T.;
RT "Cloning, expression, and characterization of stratum corneum
RT chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. BIOL. CHEM. 269:19420-19426(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE; 95314630.

```

```

RA SKYTT A., STROEMQVIST M., EGELRUD T.;
RT "Primary substrate specificity of recombinant human stratum corneum
RT chymotryptic enzyme.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995).
CC -!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
CC STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
CC SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
CC RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CC CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
CC 1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
CC OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
CC EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
CC ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L33404; G532504; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; 1DPO.
KW HYDROLASE; SERINE PROTEASE; ZMOGEN; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 22
FT PROPEP 23 29
FT CHAIN 30 253
FT ACT_SITE 70 70
FT ACT_SITE 112 112
FT ACT_SITE 205 205
FT DISULFID 36 137
FT DISULFID 55 71
FT DISULFID 144 211
FT DISULFID 176 190
FT DISULFID 201 226
FT CARBOHYD 246 246
SQ SEQUENCE 253 AA; 27525 MW; 07FDB9F7 CRC32;

```

alignment_scores:

Quality: 274.50 Length: 143
Ratio: 2.745 Gaps: 3
Percent Similarity: 69.930 Percent Identity: 38.462

alignment_block:

US-09-030-606-177 x SCCE_HUMAN ..

Align seg 1/1 to: SCCE_HUMAN from: 1 to: 253

```

7 TCGAGCCCTGGCAGCGCGGCTGTCATGGAAGAATGTCTGCTC 56
|||||
40 SerHisProTprGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG1 56
57 GGGGCTCCTGTGTCATCCGAGTGGGTGCTGTACGCCACACTGTTCC 106
|||||
56 YGlyValLeuValAsnGlnLuarGtrPvalLeuThrAlaAlaHisCysLysM 73
107 AGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGGACCAA 156
|||||
73 etAsnGlnTyrThrValHisLeuGlySerAspThrLeu..GlyAspArg 88
157 GAGCCAGGAGCAGATGTGGAGCGCCAGCCTCTCCGTACGGCACCAGCA 206
|||||
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG1 103
207 GTACAACAGACCCCTGCTGCTAAGCACTCATGCTCATCAAGTTGGAGC 256

```


FT VARIANT 186 186 E -> K.
SQ SEQUENCE 262 AA; 28889 MW; 7B954AA7 CRC32;

alignment_scores:
Quality: 272.50 Length: 155
Ratio: 2.698 Gaps: 4
Percent Similarity: 65.161 Percent Identity: 38.710

alignment_block:
US-09-030-606-177 x KLK1_HUMAN ..

Align seg 1/1 to: KLK1_HUMAN from: 1 to: 262

```
4 CACTCGAGCCCTGGCAGCGCGCACTGTCATGGAACGAATGTTCTG 53  
|||||  
34 HisSerGlnProTprGlnAlaAlaLeuTyrHisPheSerThrPheGlnCy 50  
54 CTGGGGCTCCTGTCATCCGAGTGGGTGCTGTACGGCAGACTGTT 103  
|:::|||||  
50 sGlyGlyLeuValHisArgGlnTprValLeuThrAlaAlaHisCysI 67  
104 TCCAGAACTCCTACACCATCGGCTGGGCTGCACAGCTTGAGCCGAC 153  
:::|||||  
67 LeSerAspAsnTyrGlnLeuTyrLeuGlyArgHisAsnLeu...PheAsp 82  
154 CAAGAGCCAGGAGGAGCCAGATGGTGGAGCCAGCCCTCCGTACGGCACC 203  
:::|||||  
83 AspGlnAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr 99  
204 AGAGTACACAGACCCTGCTCGCTAAC..... 231  
|:::|||||  
99 OGlyPheAsnMetSerLeuLeuGlnAsnHisThrArgGlnAlaAspGluA 116  
232 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAG 270  
|||||  
116 sPtyrSerHisAspLeuMetLeuValArgLeuThrGlnProAlaAspThr 132  
271 ...TCTGACACCATCCGAGAGCATCAGCATTTGCTTGCAGTGCCTACCGC 317  
:::|||||  
133 IleThrAspAlaValIleValIleValIleLeuProThrGlnGluProGluVa 149  
318 GGGAACTTCTGCTGCTTCTGCTGGGCTGCTGCTGCGC..... 357  
:::|||||  
149 IglySerThrCysLeuAlaSerGlyTyrGlySerIleGluProGluAsnPr 166  
358 .....AACGATGCTGTGATGCCATCCAGTCCAGACTGTGGAGGC 399  
|||  
166 HeSerPheProAspAspLeuGlnCysValAspLeuLysIleLeuProAsn 182  
400 TGGGAGTGTGAGAAG 414  
|||||  
183 AspGlnCysGlnLys 187
```

seq_name: SwissProt_37:KLK1_PAPHA

seq_documentation_block:

ID KLK1_PAPHA STANDARD; PRT; 258 AA.
AC Q28773;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN).
GN KLK1.
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CAVIARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA PERELYGINA L.M., KAMMERER C.M., HENKEL R.D.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L43121; G871814; -
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00752; 2KAI.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS;
KW ZYMOGEN; SIGNAL.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 258 GLANDULAR KALLIKREIN 1.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 117 117 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 170 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 149 216 BY SIMILARITY.
FT DISULFID 181 195 BY SIMILARITY.
FT DISULFID 206 231 BY SIMILARITY.
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
FT CARBOHYD 101 101 POTENTIAL.
FT CARBOHYD 105 105 POTENTIAL.
FT CARBOHYD 161 161 POTENTIAL.
SQ SEQUENCE 258 AA; 28339 MW; 0C5E883F CRC32;

alignment_scores:
Quality: 270.50 Length: 159
Ratio: 2.847 Gaps: 3
Percent Similarity: 59.748 Percent Identity: 37.107

alignment_block:
US-09-030-606-177 x KLK1_PAPHA ..

Align seg 1/1 to: KLK1_PAPHA from: 1 to: 258

```
7 TCGAGCCCTGGCAGCGCGCACTGTCATGGAACGAATGTTCTGCTC 56  
|||||  
32 SerGlnProTprGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysG1 48  
57 GGGCGTCTGTCATCCGAGTGGGTGCTGTACGGCAGACTGTTTCC 106  
:::|||||  
48 yGlyIleLeuValHisProGlnTprValLeuThrAlaAlaHisCysIleG 65  
107 AGAACTCCTACACCATCGGCTGGGCTGCACAGCTTGAGCCGACCAA 156  
:::|||||  
65 LyAspAsnTyrGlnLeuTyrLeuGlyArgHisAsnLeu...PheAspAsp 80  
157 GAGCAGGAGGAGCCAGATGGTGGAGCCAGCCCTCCGTACGGCAGCAGA 206  
|||  
81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97  
207 GTACAACAGACCCTGCTCGCTAAC..... 231  
:::|||||  
97 spHeAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspGluAspT 114
```



```

232 .....GACCTATGCTCATCAAGTTGGACGAATCGTGCCGAGTCT 273
      |||||
114 yserHisAspleuMeleuLeuArgLeuThrGlnProAlaGluLeuThr 130
      |||
274 GACACCATCCGAGACATCAGCATGTGCTCGAGTGCCTACCGGGGAA 323
      |||
131 AspaLaValGlnValGlnLeuProThrGlnGluProGluValGlySe 147
      |||
324 CTCTGCCTCGTTCGCTGGGT..... 348
      |||
147 rThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPheserT 164
      |||
349 .....CTGCTGGCGAAGCATGCT 366
      |||
164 yPrAspAspleuGlnCysValAspleuLysIleLeuProAsnAspLys 180
      |||
367 GTGATTGCCATCCAGTCCAGACTGTG 393
      |||
181 CysAlaLysAlaHisThrGlnLysVal 189

```

seq_name: SwissProt_37:PROS_HUMAN

seq_documentation_block:

```

ID PROS_HUMAN STANDARD: PRT: 261 AA.
AC P07288;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROSTATE SPECIFIC ANTIGEN PRECURSOR (EC 3.4.21.77) (PSA) (GAMMA-
DE SEMINOPROTEIN) (SEMINOGELASE) (SEMININ) (P-30 ANTIGEN).
GN APS.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RX MEDLINE; 89183632.
RA DIGBY M.R., ZHANG X.Y., RICHARD R.L.;
RT "Human prostate specific antigen (PSA) gene: structure and linkage to
RT the kallikrein-like gene, hK1-1."
RL NUCLEIC ACIDS RES. 17:2137-2137(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89282407.
RA KLOBECK H.-G., COMBRIATO G., SCHULZ P., ARBUSOW V., FITTLER F.;
RT "Genomic sequence of human prostate specific antigen (PSA).";
RL NUCLEIC ACIDS RES. 17:3981-3981(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-LEUKOCYTE;
RX MEDLINE; 89302090.
RA LUNDWALL A.;
RT "Characterization of the gene for prostate-specific antigen, a human
RT glandular kallikrein."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 161:1151-1159(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RX MEDLINE; 87190978.
RA LUNDWALL A., LILJA H.;
RT "Molecular cloning of human prostate specific antigen cDNA."
RL FEBS LETT. 214:317-322(1987).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RX MEDLINE; 89246551.
RA HENTTU P., VIHKO P.;
RT "cDNA coding for the entire human prostate specific antigen shows
RT high homologies to the human tissue kallikrein genes."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 160:903-910(1989).
RN [6]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE-PROSTATE;
RX MEDLINE; 89165891.
RA RIEGMAN P.H.J., VIETSTRA R.J., VAN DER KORPUT J.A.G.M., ROMJON J.C.,
RA TRAPMAN J.;
RT "Characterization of the prostate-specific antigen gene: a novel
RT human kallikrein-like gene."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 159:95-102(1989).
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE-PROSTATE;
RA MONNE M.M., MORENO J.M., MELE C.M., MULHOLLAND G.M., GOMELLA L.G.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [8]
RP SEQUENCE OF 17-261 FROM N.A.
RX MEDLINE; 88289366.
RA SCHULZ P., STUCKA R., FELDMANN H., COMBRIATO G., KLOBECK H.-G.,
RA FITTLER F.;
RT "Sequence of a cDNA clone encompassing the complete mature human
RT prostate specific antigen (PSA) and an unspliced leader sequence."
RL NUCLEIC ACIDS RES. 16:6226-6226(1988).
RN [9]
RP SEQUENCE OF 25-261.
RX MEDLINE; 86205857.
RA WATT K.W.K., LEE P.J., M'TIMKULU T., CHAN W.P., LOOR R.;
RT "Human prostate-specific antigen: structural and functional
RT similarity with serine proteases."
RL PROC. NATL. ACAD. SCI. U.S.A. 83:3166-3170(1986).
RN [11]
RP SEQUENCE OF 25-261.
RX MEDLINE; 88082806.
RA SCHALLER J., AKIYAMA K., TSUDA R., HARA M., MARTI T., RICKLI E.E.;
RT "Isolation, characterization and amino-acid sequence of gamma-
RT seminoprotein, a glycoprotein from human seminal plasma."
RL EUR. J. BIOCHEM. 170:111-120(1987).
RN [11]
RP 3D-STRUCTURE MODELLING.
RX MEDLINE; 95218633.
RA VILLOUTREIX B.O., GETZOFF E.D., GRIFFIN J.H.;
RT "A structural model for the prostate disease marker, human prostate-
RT specific antigen."
RL PROTEIN SCI. 3:2033-2044(1994).
CC -1- FUNCTION: PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS SEMINAL
CC VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF THE SEMINAL
CC COAGULUM.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: TYR-1-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14810; G296671; -
DR EMBL; X13940; G35723; -
DR EMBL; X13941; E14018; ALT_SEQ.
DR EMBL; X13942; E14019; -
DR EMBL; X13943; E14020; -
DR EMBL; X13944; E14022; -
DR EMBL; X05332; G35741; -
DR EMBL; X07730; -; NOT_ANNOTATED_CDS.
DR EMBL; M27274; G190553; -
DR EMBL; M26663; G618464; -
DR EMBL; U17040; G595946; -
DR EMBL; M24543; G511857; ALT_SEQ.
DR PIR; S03604; S03604.
DR PIR; A26757; A26757.
DR PIR; A32297; A32297.
DR PIR; A32423; A32423.
DR PDB; 1PFA; 26-JAN-95.

```



```

DR MIM; 176820;
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ANTIGEN; ZYMOGEN; SIGNAL;
3D-STRUCTURE.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 24 ACTIVATION PEPTIDE.
FT CHAIN 25 261 PROSTATE SPECIFIC ANTIGEN.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 69 69
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CONFLICT 64 64 A -> T (IN REF. 8).
FT CONFLICT 73 73 I -> T (IN REF. 5).
FT CONFLICT 86 86 T -> I (IN REF. 5).
FT CONFLICT 94 94 H -> T (IN REF. 9).
FT CONFLICT 136 136 V -> M (IN REF. 8).
FT CONFLICT 165 168 FLTP -> HLYDQM (IN REF. 9).
FT CONFLICT 175 175 D -> Q (IN REF. 9).
FT CONFLICT 175 175 D -> P (IN REF. 5).
FT CONFLICT 184 184 C -> Q (IN REF. 5).
FT CONFLICT 260 260 N -> D (IN REF. 5).
SO SEQUENCE 261 AA; 28741 MM; 4723B52E CRC32;

```

```

alignment_scores:
    Quality: 270.00      Length: 170
    Ratio: 2.700        Gaps: 4
    Percent Similarity: 58.824    Percent Identity: 34.706

```

alignment_block:
US-09-030-606-177 x PROS_HUMAN ..

Align seg 1/1 to: PROS_HUMAN from: 1 to: 261

```

4 CACTCGCAGCCCTGGCAGCGCGGCTGTCATGGAACAATGTTCTG 53
|||||
34 HisSerGlnProTrrGlnValLeuValAlaSerArgLysAlaValCy 50
54 CTGGGGCTGCTGTCATCCGACATGGGTGTCAGCCGACACTGTT 103
|:::|
50 sglyglyValleuValHisProGlnTrrValleuThrAlaAlaHisCysI 67
104 TCCAGAACTTCCTACACCATCGGGCTGGCCTGCACAGTCTTGAGCCGAC 153
:::|
67 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 83
154 CAAGAGCCAGGAGGACCATGATGTTGAGGCCAGCCTCTCCGTACGGCACC 203
:::|
84 GluAspThrCly...GlnValPheGlnValSerHisSerPheProHisPr 99
204 AGAGTACACAGACCCCTGCTCGCTAAC..... 231
|
99 OLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspa 116
232 .....GACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAG 270
|||||
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 132
271 TCTGACACCATCCGAGCATGCTTGCAGTGCCTTACCGCGGG 320
:::|
133 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGl 149
321 GAAGTCTGCTGCTGTTCTGCTGGGTT..... 348
|
149 YThrThrCysTyrAlaSerGlyTrrPglySerIleGluProGluGluPheL 166

```

```

349 .....CTGCTGGCGAAGCAT 363
:::|
166 eutThrProLysLysLeuGlnCysValAspLeuHisValIleSerAsnAsp 182
364 GCTGTGATTCGCATCCAGTCCAGACTGTG..... 393
:::|
183 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 199
394 .GGAGGCTGG 402
|||
199 aglyArgTrr 202

```

seq_name: SwissProt_37:KLK2_HUMAN

seq_documentation_block:

```

ID KLK2_HUMAN STANDARD; PRT; 261 AA.
AC P20151;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 2 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (PROSTATE) (HGK-1).
GN KLK2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88054467.
RA SCHEDLICH L.J., BENNETTS B.H., MORRIS B.J.;
RT "Primary structure of a human glandular kallikrein gene.";
RL DNA 6:429-437(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-PROSTATE;
RX MEDLINE; 92324494.
RA RIEGMAN P.H., VLIETSTRA R.J., DER KOPPUT H.A., ROMIJN J.C.,
RA TRAPMAN J.;
RT "Identification and androgen-regulated expression of two major human
RT glandular kallikrein-1 (hgk-1) mRNA species.";
RL MOL. CELL. ENDOCRINOL. 76:181-190(1991).
CC -|- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC -|- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18156; -; NOT_ANNOTATED_CDS.
DR EMBL; M18157; G386842; -.
DR EMBL; S39329; E79223; -.
DR PIR; A29586; A29586.
DR MIM; 147960; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P07288; 1PFA.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZYMOGEN; SIGNAL.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 261 GLANDULAR KALLIKREIN 2.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.

```


FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 102 102 POTENTIAL.
SQ SEQUENCE 261 AA; 28671 MW; 99CF6275 CRC32;

alignment_scores:

Quality: 266.50 Length: 160
Ratio: 2.835 Gaps: 3
Percent Similarity: 58.750 Percent Identity: 36.250

alignment_block:

US-09-030-606-177 x KIK2_HUMAN ..

Align seg 1/1 to: KIK2_HUMAN from: 1 to: 261

4 CACTCGAGCCCTGGCAGCGCGGCACTGTCATGGAAGAATGTTCTG 53
|||||
34 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCys 50
54 CTCGGGCGTCTGTCATCCGAGTGGGTGTCAGCCGACACTGTT 103
|::|
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTGAGCGCGAC 153
:::|
67 eulysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 83
154 CAAGAGCCAGGAGCCAGATGTTGGAGGCCACCTCTCCGTACGGCACCC 203
:::|
84 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 99
204 AGAGTACAACAGACCTTGCTC..... 225
|
99 oleuTyraSmetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 116
226GCTAAGACCTCATGTCATCAAGTGGAGCAATCCGTCCGAG 270
:::|
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 132
271 TCTGACACCATCCGAGACATCAGCATTCCTGCGAGTCCCTACCGCGG 320
::|
133 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 149
321 GAACCTCTGCTGCTTCTGCTGGGGT..... 348
|:::|
149 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 166
349CTGCTGGCGAAGCAT 363
|
166 euArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAsp 182
364 GCTGTGATGGCATCCAGTCCCGAGACTGTG 393
|
183 MetCysAlaArgAlaTyrSerGluLysVal 192

seq_name: SwissProt_37:PROS_MACMU

seq_documentation_block:

ID PROS_MACMU STANDARD; PRT; 261 AA.
AC P33619;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROSTATE SPECIFIC ANTIGEN PRECURSOR (EC 3.4.21.35) (PSA) (GAMMA-
DE SEMINOPROTEIN).
GN APS.
OS MACACA MULATTA (RHESUS MACAQUE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93363642.
RA GAUTHIER E.R., CHAPDELAINE P., TREMBLAY R.R., DUBE J.Y.;
RT "Characterization of rhesus monkey prostate specific antigen cDNA.";
RT BIOCHIM. BIOPHYS. ACTA 1174:207-210(1993).
CC -1- FUNCTION: PSA PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS
CC SEMINAL VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF
CC THE SEMINAL COAGULUM. PSA IS A GLANDULAR KALLIKREIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X73560; G311844; -.
DR PIR; S34239; S34239.
DR PIR; S35711; S35711.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSP; P07288; 1DPA.
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ANTIGEN; ZYMOGEN; SIGNAL.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 24 ACTIVATION PEPTIDE.
FT CHAIN 25 261 PROSTATE SPECIFIC ANTIGEN.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 102 102 POTENTIAL.
SQ SEQUENCE 261 AA; 28816 MW; 0B73118F CRC32;

alignment_scores:

Quality: 262.00 Length: 170
Ratio: 2.646 Gaps: 4
Percent Similarity: 58.235 Percent Identity: 33.529

alignment_block:

US-09-030-606-177 x PROS_MACMU ..

Align seg 1/1 to: PROS_MACMU from: 1 to: 261

4 CACTCGAGCCCTGGCAGCGCGGCACTGTCATGGAAGAATGTTCTG 53
|||||
34 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCys 50
54 CTCGGGCGTCTGTCATCCGAGTGGGTGTCAGCCGACACTGTT 103
|::|
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTGAGCGCGAC 153
:::|
67 leArgSerAsnSerValIleLeuLeuGlyArgHisAsnProTyrTyrPro 83
154 CAAGAGCCAGGAGCCAGATGTTGGAGGCCACCTCTCCGTACGGCACCC 203
:::|
84 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 99
204 AGAGTACAACAGACCTTGCTGCTAC..... 231
|
|

99 oleuTyraSmetSerleuLeuLysAsnArgTyrleuGlyProGlyAspa 116
232GACCTCATGCTCATCACTGACGACATCCGTCCGAG 270
116 spSerSerHisAspLeuMetleuLeuArgleuSerGluProAlaGluIle 132
271 TCTGACACCATCCGAGACATCAGCATGCTTCCGAGTGCCTACCGCGG 320
133 ThrAspAlaValaGlnValaLeuAspLeuProThrTrpGluProGluLeuG 149
321 GAACCTCTGCTGCTTCTGCTGGGT..... 348
149 YThrThrCysTyrAlaSerGlyTyrGlySerIleGluProGluGluHisL 166
349CTGCTGGCGAAGCAT 363
166 eutThrProLysLysleuGlnCysValaSpluHisIleIleSerAsnAsp 182
364 GCTGTGATGTCATCCAGTCCAGACTGTG..... 393
183 ValCysAlaGlnValaHisSerGlnLysValaThrLysPheMetleuCysAl 199
394 .GGAGCGCTGG 402
199 aglySerTrp 202
seq_name: SwissProt_37:KLKL_MOUSE
seq_documentation_block:
ID KLKL_MOUSE STANDARD; PRT; 259 AA.
AC P15948;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (MGK-22) (EPIDERMAL GROWTH FACTOR-BINDING PROTEIN TYPE A) (EGF-BP A)
DE (NERVE GROWTH FACTOR BETA CHAIN ENDOPEPTIDASE) (BETA-NGF-
DE ENDOPEPTIDASE).
GN KLK22 OR KLK-22.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-SALIVARY GLAND;
RX MEDLINE: 88107594.
RA DRINKWATER C.C., EVANS B.A., RICHARDS R.I.;
RT "mouse glandular kallikrein genes: identification and
RT characterization of the genes encoding the epidermal growth factor
RT binding proteins.";
RL BIOCHEMISTRY 26:6750-6756(1987).
RN [2]
RP SEQUENCE OF 25-54.
RX MEDLINE: 91190897.
RA FAHNESTOCK M., WOO J.E., LOPEZ G.A., SNOW J., WALZ D.A., ARICI M.J.,
RA MOBLEY W.C.;
RT "Beta-NGF-endopeptidase: structure and activity of a kallikrein
RT encoded by the gene mgk-22.";
RL BIOCHEMISTRY 30:3443-3450(1991).
RN [3]
RP SEQUENCE OF 25-41.
RC TISSUE-SUBMANDIBULAR GLAND;
RX MEDLINE: 92348355.
RA PETERS J., TAKAHASHI S., TADA M., MIYAKE Y.;
RT "mgk-6-derived true tissue kallikrein is synthesized, processed, and
RT targeted through a regulated secretory pathway in mouse pituitary
RT AtT-20 cells.";
RL J. BIOCHEM. 111:643-648(1992).
RN [4]
RP SEQUENCE OF 17-54 AND 70-120 FROM N.A.
RX MEDLINE: 87250386.
RA EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;
RT "mouse glandular kallikrein genes. structure and partial sequence

RT analysis of the kallikrein gene locus.";
RL J. BIOL. CHEM. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17979; G387167; ALT_SEQ.
DR EMBL: M17977; G387167; JOINED.
DR EMBL: M17978; G387167; JOINED.
DR EMBL: M18598; G198560; -.
DR EMBL: M18618; G198561; -.
DR PIR: A29746; A29746.
DR PIR: A38356; A38356.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PFAM: PF00089; trypsin; 1.
DR HSSP: P00757; 1SGF.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZYMOGEN; SIGNAL.
KM
FT SIGNAL 1 17
FT PROPEP 18 24 ACTIVATION PEPTIDE.
FT CHAIN 25 259 GLANDULAR KALLIKREIN K22.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
FT ACT_SITE 211 211 CHARGE RELAY SYSTEM.
FT DISULFID 31 171 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 150 217 BY SIMILARITY.
FT DISULFID 182 196 BY SIMILARITY.
FT DISULFID 207 232 BY SIMILARITY.
FT CARBOHYD 102 102 PROBABLE.
SQ SEQUENCE 259 AA; 28384 MW; 3B02165A CRC32;
alignment_scores: Quality: 260.00 Length: 180
Ratio: 2.342 Gaps: 6
Percent Similarity: 61.667 Percent Identity: 33.889
alignment_block:
US-09-030-606-177 x KLKL_MOUSE ..
Align seg 1/1 to: KLKL_MOUSE from: 1 to: 259
4 CACTCGAGCCCTGGAGCGCGCACTGTCATMGAAAGCAATGTCTCTG 53
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
34 AsnSerGlnProTrpGlnValaValaValaTyrTyrLeuAspGluTyrLeuG 50
54 CTCGGCGCTCTGTGATCCGCACTGGGTGCTGTACGCCGACACTGT 103
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
50 sGlyGlyValleuLeuAspArgAsnTrpValleuThrAlaAlaHisCysT 67
104 TCCGAACCTCTACACCATCGGCGCTGGCGCTGCACAGTCTTGAGGCGGAC 153
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
67 yrgluAspLysTyrAsnIleTrpLeuGlyLysAsnLysleuPheGlnAsp 83
154 CAAGAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGCCACC 203
|||||:::|||||:::|||||:::|||||:::|||||
84 ...GluProSerAlaGlnHisArgleuValSerLysSerPheProHisPr 99


```

67 yrAspAspAsnTyrLysValTrpLeuGlyLysAsnAsnLeuPheLysAsp 83
154 CAAGAGCCAGGAGGAGCCAGATGCTGAGAGCCAGCCCTCTCCGTACGGACCC 203
      |||||:::|||||:::|||||:::|||||
84 ...GluProSerAlaGlnHisArgPheValSerLysAlaIleProHisPr 99
204 AGAGTACACAGAGACCCTGCTC..... 225
      |:::|||||:::|||||
99 oGlyPheAsnMetSerLeuMetArgLysHisIleArgPheLeuGluTyrA 116
226 .....GCTAACGACCTCATGCTCATCAAGTTGAGCGAATCCGTCCGAG 270
      :::|||||:::|||||:::|||||:::|||||
116 sPtyrSerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIle 132
271 TCTGACACCATCCGAGCATCAGCATGCTTCCAGTCCCTACCGCGGG 320
      :::|||||:::|||||:::|||||:::|||||
133 ThrAspThrValLysProIleThrLeuProThrGluGluProLysLeuGl 149
321 GAACCTCTGCCTCGTTCTGCTGGGCTGCTGCGG..... 357
      |:::|||||:::|||||:::|||||
149 ySerThrCysLeuAlaSerGlyTrpGlySerIleThrProThrLysPheG 166
358 .....AACGATGCTGTGATGTCATCCAGTCCAGACTGTGGAGGCTGG 402
      :::|||||:::|||||:::|||||:::|||||
166 lnhethrAspAspLeuTyrCysValAsnLeuLysLeuLeuProAsnGlu 182
403 GAGTGTGAGAAG 414
      :::|||||
183 AspCysAlaLys 186

```

seq_name: SwissProt_37:KLKA_RAT

seq_documentation_block:

ID KLKA_RAT STANDARD; PRT; 244 AA.

AC P36375;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE GLANDULAR KALLIKREIN 10 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)

DE (T-KININOGENASE) (K10) (PROTEINASE B) (ENDOPEPTIDASE K) (FRAGMENT).

GN KLK10 OR KLK-10.

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MORIDAE; MURINAE; RATTUS.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE-SUBMANDIBULAR GLAND, AND KIDNEY;

RX MEDLINE; 93041794.

RA MA J.-X., CHAO J., CHAO L.;

RT "Molecular cloning and characterization of rklk10, a cDNA encoding T-

kininogenase from rat submandibular gland and kidney.";

RL BIOCHEMISTRY 31:10922-10928(1992).

RN [2]

RP SEQUENCE OF 10-32; 95-124 AND 179-232.

RX MEDLINE; 91224135.

RA GUTMAN N., ELMOUJAHED A., BRILLARD M., DU SORBIER B., GAUTHIER F.;

RT "Microheterogeneity of rat submaxillary gland kallikrein k10, a

member of the kallikrein family.";

RL EUR. J. BIOCHEM. 197:425-429(1991).

RN [3]

RP SEQUENCE OF 10-32 AND 97-133.

RC TISSUE-SUBMAXILLARY GLAND;

RX MEDLINE; 90153911.

RA XIONG W., CHEN L.-M., CHAO J.;

RT "Purification and characterization of a kallikrein-like

T-kininogenase.";

RL J. BIOL. CHEM. 265:2822-2827(1990).

RN [4]

RP SEQUENCE OF 10-32 AND 97-117.

RX MEDLINE; 88198057.

RA KATO H., NAKANISHI E., ENJOJOI K., HAYASHI I., OH-ISHI S., IWANAGA S.;

RT "Characterization of serine proteinases isolated from rat

submaxillary gland: with special reference to the degradation of rat

```

RT kininogens by these enzymes.";
RL J. BIOCHEM. 102:1389-1404(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS PROTEIN MAY BE
CC INVOLVED IN THE REGULATION OF RENAL FUNCTION.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED
CC BY A DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND, WHERE IT IS
CC FOUND IN THE GRANULAR CONVOLUTED TUBULE AND STRIATED DUCT CELLS.
CC IT IS LIKELY THAT THE ENZYME IS MAINLY SYNTHESIZED IN THE GRANULAR
CC CONVOLUTED TUBULES AND THEN TRANSFERRED TO OTHER TISSUES BY
CC RELEASE INTO THE VASCULATURE OR INTERSTITIAL SPACE.
CC -1- PTM: PROBABLY N- AND O-GLYCOSYLATED. IT HAS CARBOHYDRATE MOTIFIES
CC OF ALPHA-METHYL-D-MANNOSIDE AND N-ACETYL-D-GLUCOSAMINE GROUPS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; S48142; G259431; -.
DR PIR; A35545; A35545.
DR PIR; B35545; B35545.
DR PIR; A44284; A44284.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00759; 1TON.
DR KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; SIGNAL;
KW MULTIGENE FAMILY.
FT FT 1 1
FT SIGNAL <1 3
FT PROPEP 4 9
FT CHAIN 10 244
FT CHAIN 10 96
FT CHAIN 97 244
FT ACT_SITE 48 48
FT ACT_SITE 103 103
FT ACT_SITE 196 196
FT DISULFID 16 156
FT DISULFID 33 49
FT DISULFID 135 202
FT DISULFID 167 181
FT DISULFID 192 217
FT CARBOHYD 76 76
FT CARBOHYD 91 91
FT CARBOHYD 126 126
FT CARBOHYD 142 142
FT CONFLICT 29 29
FT CONFLICT 115 116
FT CONFLICT 128 128
FT CONFLICT 133 133
SQ SEQUENCE 244 AA; 27305 MW; 7661FB5F CRC32;

```

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 253.50 | length: | 161 |
| Ratio: | 2.613 | Gaps: | 5 |
| Percent Similarity: | 60.248 | Percent Identity: | 35.404 |

alignment_block:

US-09-030-606-177 x KLKA_RAT ..

Align seg 1/1 to: KLKA_RAT from: 1 to: 244

[illegible]

```

seq_name: Swissprot_37:KLK_MOUSE
seq_documentation_block:
ID   KLK_MOUSE      STANDARD;          PRT;          261 AA.
AC   P15946;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GLANDULAR KALLIKREIN K11 PRECURSOR (EC 3.4.21.35), SUBMANDIBULAR
DE   (TISSUE KALLIKREIN) (MGK-11).
GN   KLK11 OR KLK-11.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BALB/C; TISSUE=LIVER;
RX   MEDLINE; 89083511.
RA   DRINKWATER C.C., RICHARDS R.I.;
RT   "Sequence of mgk-11, a mouse glandular kallikrein gene.";
RL   NUCLEIC ACIDS RES. 16:10918-10918(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85257431.
RA   EVANS B.A., RICHARDS R.I.;
RT   "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT   are contiguous.";
RL   EMBO J. 4:133-138(1985);
RN   [3]
RP   SEQUENCE OF 16-54 AND 69-122 FROM N.A.
RX   MEDLINE; 87250386.
RA   EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;

```

```

RT "Mouse glandular kallikrein genes. Structure and partial sequence
RT analysis of the kallikrein gene locus.";
RL J. BIOL. CHEM. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

| CC | Sequence | Mass | Charge | Protein |
|----------|---|------|--------|--------------------------------|
| DR | EMBL: X13215; G53066; - | 261 | 1 | PROBABLE. |
| DR | EMBL: X13216; G582395; ALT_SEQ. | 261 | 1 | ACTIVATION PEPTIDE (PROBABLE). |
| DR | EMBL: X13217; E14069; - | 261 | 1 | GLANDULAR KALLIKREIN K11. |
| DR | EMBL: X13217; E14070; - | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | EMBL: X13218; E7100; - | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | EMBL: M18590; G554179; ALT_SEQ. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | EMBL: M18610; G198537; - | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | PIR: S01971; S01971. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | MGI: 892023; K1K11. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | PROSITE: PS00134; TRYPsin_HIS; 1. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | PROSITE: PS00135; TRYPsin_SER; 1. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | PFAM: PF00089; trypsin; 1. | 261 | 1 | CHARGE RELAY SYSTEM. |
| DR | HSSP: P00757; 1SGF. | 261 | 1 | CHARGE RELAY SYSTEM. |
| KW | HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; | 261 | 1 | CHARGE RELAY SYSTEM. |
| KW | MULTIGENE FAMILY; ZYMOGEN; SIGNAL. | 261 | 1 | CHARGE RELAY SYSTEM. |
| FT | SIGNAL | 1 | 18 | PROBABLE. |
| FT | PROPEP | 19 | 24 | ACTIVATION PEPTIDE (PROBABLE). |
| FT | CHAIN | 25 | 261 | GLANDULAR KALLIKREIN K11. |
| FT | ACT_SITE | 65 | 65 | CHARGE RELAY SYSTEM. |
| FT | ACT_SITE | 120 | 120 | CHARGE RELAY SYSTEM. |
| FT | ACT_SITE | 213 | 213 | CHARGE RELAY SYSTEM. |
| FT | DISULFID | 31 | 173 | CHARGE RELAY SYSTEM. |
| FT | DISULFID | 50 | 66 | CHARGE RELAY SYSTEM. |
| FT | DISULFID | 152 | 219 | CHARGE RELAY SYSTEM. |
| FT | DISULFID | 184 | 198 | CHARGE RELAY SYSTEM. |
| FT | DISULFID | 209 | 234 | CHARGE RELAY SYSTEM. |
| FT | CARBOHYD | 102 | 102 | CHARGE RELAY SYSTEM. |
| SEQUENCE | 261 AA; 28727 MW; FD904C22 CRC32; | | | |

```
alignment_scores:
  Quality: 250.50
  Ratio: 2.665
  Percent Similarity: 64.384
  Length: 146
  Gaps: 3
  Percent Identity: 36.301
```

alignment_block:

US-09-030-606-177 x KLKA_MOUSE

Align seg 1/1 to: KKA_MOUSE from: 1 to: 261

```

4  CACTCCAGCCCTGGCAGGCGGCACTGGTCATGGAADAAGCAATGTCTG 53
   ::::::::::::::::::::|:::|
34 AsnSerGlnProTrpHisValAlaValTyrArgTyrAsnLysTyrIleCy 50
   ::::::::::::::::::::|:::|
54 CTCGGCGCTCTGTGCATCCGCAATGGGTCTGTGCAGCCGACACTGTT 103
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
50 sGlyGlyValLeuLeuAspArgAsnTrpValLeuThrAlaIleHisCysH 67
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
104 TCACAAGTCTCTACACCAATCGGCTGGCTGCACAGTCTTGAGCGGAC 153
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
67 AsValSerGlnTyrAsnValTrpLeuGlyLysThrLysLeu...PheGln 82
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
154 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTTCTCCGTACGGCACC 203

```



```

      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
83 ArgGluProSerAlaGlnHisArgMetValSerLysSerPheProHisPr 99
204 AGAGTACACAGACAGCCCTGCTC..... 225
99 OAsPTYrAsnMetSerLeuLeuIleIleHisAsnProGluProGluAspa 116
226 .....GCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAG 270
116 spcIuserAsnAspLeuMetLeuLeuArgLeuSerGluProAlaAspIle 132
271 TCTGACACCATCCGAGACATGATGCTTCGACAGTGCCTACCGCGG 320
133 ThrAspAlaValLysProIleAlaLeuProThrGluGluProLysLeuG1 149
321 GAACCTCTGCCTCGTTCTGCTGGGGTCTGCTGGCGACGATGCTGTGA 370
149 ySerThrCysLeuValSerGlyTyrGly.....SerIleT 161
371 TTGCCATCCAGTCCAGACTGTGGGAGGCTGGAGTGT 408
161 hrProThrLysPheGlnThrProAspAspLeuGlnCys 173

seq_name: SwissProt_37:TRY3_CHICK

seq_documentation_block:
ID TRY3_CHICK STANDARD; PRT; 248 AA.
AC Q90629;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN II-P29 PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RX MEDLINE; 95251611.
RA WANG K., GAN L., LEE I., HOOD L.E.;
RT "Isolation and characterization of the chicken trypsinogen gene
family."
RL BIOCHEM. J. 307:471-479(1995).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U15157; G603907; -
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSP; P00763; IDPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 16 BY SIMILARITY.
FT PROPEP 17 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPSIN II-P29.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.

```

```

FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 248 AA; 26622 MW; F7FE97E9 CRC32;

alignment_scores:
Quality: 249.00 Length: 143
Ratio: 2.465 Gaps: 4
Percent Similarity: 70.629 Percent Identity: 37.063

alignment_block:
US-09-030-606-177 x TRY3_CHICK ..

Align seg 1/1 to: TRY3_CHICK from: 1 to: 248

4 CACTCGACAGCCCTGGCAGCGGCGCACTGTGTCATGGAACGAATGTCTTG 53
||||| |||::|||::|||::|||::|||::|||::|||::|||::|||
35 HisSerValProTyrGlnValSerLeu..AsnSerGlyTyrHisPheCys 50
54 CTGCGGCGCTCTGTGTCATCCGACAGTGGGTGCTGTACGCCGACACTGT 103
|||::||| |||::|||::|||::|||::|||::|||::|||::|||::|||
50 sGlyGlySerLeuIleAsnSerGlnTyrValLeuSerAlaAlaHisCysT 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCGCGAC 153
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
67 yrlYSerArgIleGlnValArgLeuGlyGlyTyrAsnIleAspVal... 82
154 CAAGAGCCAGGAGCCAGATGTGTGAGGCCACCTCTCCGTACGCCACCC 203
||||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|
83 GlnGluAspSerGluValAlaArgSerSerSerValIleIleArgHisPr 99
204 AGAGTACACAGACACCTTCTGCTGCTAACGACCTCATGCTCATCAAGTTG 253
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
99 OlyTyrSerSerIleThrLeuAsnAsnAspIleMetLeuIleLysLeuA 116
254 ACGAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATGTCTCG 303
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
116 LasSerAlaValGluTyrSerAlaAspIleGlnProIleAlaLeuProSer 132
304 CAGTGCCTTACCGCGGGAACCTTCTGCTGCTGCTGCTGGGT...CT 350
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
133 SerCysAlaLysAlaGlyThrGluCysLeuIleSerGlyTyrGlyAsnTh 149
351 GCTGGCGAAC.....GATGCTGTGATTGCCATCCAGTCCC 385
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
149 rLeuSerAsnGlyTyrAsnTyrProGluLeuGlnCysLeuAsnAlaP 166
386 AGACTGTGGAGGCTGGAGTGTGAGAAG 414
166 rLeuLeuSerAspGlnGlnCysGlnGlu 175

seq_name: SwissProt_37:TRY1_CANFA

seq_documentation_block:
ID TRY1_CANFA STANDARD; PRT; 246 AA.
AC P06871;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN, CATIONIC PRECURSOR (EC 3.4.21.4).
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86284628.
RA PINSKY S.D., LAForge K.S., SCHEELE G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT in the dog pancreas."

```


RX MEDLINE; 88011252.
RA FUJINAGA M., JAMES M.N.G.;
RT "Rat submaxillary gland serine protease, tonin. Structure solution
RT and refinement at 1.8-A resolution.";
RL J. MOL. BIOL. 195:373-396(1987).
CC -1- FUNCTION: THIS PROTEIN HAS BOTH TRYPSIN- AND CHYMOTRYPSIN-LIKE
CC ACTIVITIES, BEING ABLE TO RELEASE ANGIOTENSIN II FROM ANGIOTENSIN
CC I OR ANGIOTENSINOGEN.
CC -1- SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: FOUND IN SUBMAXILLARY GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11565; G205034; -.
DR EMBL; M23878; G207413; -.
DR EMBL; M23877; G207413; JOINED.
DR EMBL; M26533; G206776; -.
DR PIR; A00945; KQRTTN.
DR PIR; A30971; A30971.
DR PIR; A34050; A34050.
DR PIR; C23863; C23863.
DR PIR; B33359; B33359.
DR PIR; A32340; A32340.
DR PDB; 1TON; 16-JAN-88.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; MULTIGENE FAMILY; ZYMOGEN;
KW SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 259
FT ACT_SITE 63 63
FT ACT_SITE 118 118
FT ACT_SITE 211 211
FT DISULFID 31 171
FT DISULFID 48 64
FT DISULFID 150 217
FT DISULFID 182 196
FT DISULFID 207 232
FT CARBOHYD 106 106
FT CARBOHYD 189 189
FT STRAND 26 26
FT STRAND 29 30
FT TURN 33 34
FT TURN 37 38
FT STRAND 39 43
FT STRAND 47 54
FT TURN 55 56
FT STRAND 57 60
FT HELIX 62 64
FT STRAND 70 73
FT STRAND 77 77
FT TURN 78 79
FT TURN 83 84
FT STRAND 86 88
FT STRAND 90 95
FT TURN 97 98
FT TURN 116 117
FT STRAND 120 124
FT TURN 146 147
FT STRAND 149 154
FT STRAND 168 168
FT STRAND 170 177
FT HELIX 179 181

FT HELIX 183 186
FT TURN 188 189
FT HELIX 190 193
FT STRAND 194 198
FT TURN 200 201
FT STRAND 205 205
FT TURN 208 209
FT TURN 211 212
FT STRAND 214 217
FT TURN 218 219
FT STRAND 220 225
FT TURN 235 236
FT STRAND 239 243
FT HELIX 244 257
SQ SEQUENCE 259 AA; 28248 MW; 745DA500 CRC32;

alignment_scores: Quality: 246.50 Length: 161
 Ratio: 2.541 Gaps: 5
Percent Similarity: 60.248 Percent Identity: 34.783

alignment_block:
US-09-030-606-177 x KLK2_RAT ..

Align seg 1/1 to: KLK2_RAT from: 1 to: 259

4 CACTCGACCCCTGGCAGGCGCAGCTGTCATGGAACGAATTGTCG 53
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
34 AsnSerGlnProTyrGlnValAlaValIle.....AsnGluTyrLeucy 48
54 CTCGGCGCTCCTGCTGTCATCCGCGAGTGGGTGCTGTCACCCGACACTGT 103
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
48 sGlyGlyValLeuIleAspProSerTyrValIleThrAlaAlaHisCyst 65
104 TCCAGACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
65 yrSerAsnAsnTyrGlnValLeuLeuGlyArgAsnAsnLeuPheLysAsp 81
154 CAGAGCCGAGGAGCCAGATGTTGAGGCCAGCCCTCTCCGTACGGCACCC 203
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
82 ..GluProPheAlaGlnArgArgLeuValArgGlnSerPheArgHisPr 97
204 AGAGTACACAGACCCCTTGCTC..... 225
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
97 oAspTyr...IleProLeuIleValThrAsnAspThrGlnProValH 113
226GCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTC 267
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
113 IsAspHisSerAsnAspLeuMetLeuHisLeuSerGluProAlaAsp 129
268 GAGTCTGACACCATCCGGAGCATCAGCATTCGTCGACATGCCCTACCGC 317
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
130 IleThrGlyGlyValLysValIleAspLeuProThrLysGluProLysVa 146
318 GGGGAACTCTTGCCTCGTTTCTGGCTGGGGT..... 348
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
146 lGlySerThrCysLeuAlaSerGlyTyrGlySerThrAsnProSerGluM 163
349CTGCTGGCGAAC 360
|||||:::|||||
163 etValValSerHisAspLeuGlnCysValAsnIleHisLeuSerAsn 179
361 GATGCTGTGATTGCCATCCAGTCCAGACTGTG 393
:::|||||:::|||||
180 GluLysCysIleGluThrTyrLysAspAsnVal 190

seq_name: SwissProt_37:KLK7_RAT
seq_documentation_block:
ID KLK7_RAT STANDARD; PRT; 261 AA.
AC P36373;
DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 7, SUBMANDIBULAR/RENAL PRECURSOR (EC 3.4.21.35)
DE (TISSUE KALLIKREIN) (RGK-7) (RSKG-7) (ESTERASE B) (PROTEINASE A).
GN KLK7 OR KLK-7.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89088074.
RA CHEN Y.-P., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of two rat renal kallikrein
genes.";
RT BIOCHEMISTRY 27:7189-7196(1988).
RN [2]
RP SEQUENCE OF 25-75.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE; 88198057.
RA KATO H., NAKANISHI E., ENJOJI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT "Characterization of serine proteinases isolated from rat
submaxillary gland: with special reference to the degradation of rat
kallikrein by these enzymes.";
RT J. BIOCHEM. 102:1389-1404(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYS1-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19647; G205000; .
DR PIR; A31136; A31136.
DR PIR; B41429; B41429.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00759; 1TON.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZMOGEN; SIGNAL.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 261 GLANDULAR KALLIKREIN 7.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 108 108 POTENTIAL.
FT CONFLICT 35 35 S -> D (IN REF. 2).
FT CONFLICT 46 46 T -> S (IN REF. 2).
SQ SEQUENCE 261 AA; 28972 MW; 27B65B42 CRC32;

alignment_scores: Quality: 245.50 Length: 161
Ratio: 2.557 Gaps: 4
Percent Similarity: 59.627 Percent Identity: 34.161

alignment_block:
US-09-030-606-177 x KLK7_RAT ..
Align seg 1/1 to: KLK7_RAT from: 1 to: 261

4 CACTCGACCCCTGGCAGCGCGACCTGGTCATGAAAACGAATTGTTCTG 53
:::|||||||||||||:|||||:|||||:|||||:|||||:|||||
34 AsnSerGlnProThrGlnValAlaLeuTyrSerPheThrLysTyrLeu 50
CTCGGCGTCTGCTGCATCCGCAGTGGGTGCTGTACGCCGACACTGT 103
:|||||||||||||:|||||:|||||:|||||:|||||:|||||
50 sGlyGlyValLeuIleAspProSerTrpValIleThrAlaAlaHisCys 67
TCCAGAACTCTCAACACCATCGGGCTGGCCCTGCACAGTCTTGAGGCCGAC 153
:::|||||:|||||:|||||:|||||:|||||:|||||
67 erSerAsnAsnTyrGlnValTrpLeuGlyArgAsnAsnLeuLeuGluAsp 83
154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACC 203
|||||:|||||:|||||:|||||:|||||:|||||
84 ..GluProPheAlaGlnHisArgLeuValSerGlnSerPheProHisPr 99
204 AGAGTACAACAGACCTTGCTC..... 225
:|||||:|||||:|||||:|||||:|||||:|||||
99 AspTyr..LysProPheLeuMetArgAsnHisThrArgLysProGlyA 115
226GCTAACGACCTCATGCTCATCAAGTTGACGAAATCCGTGTC 267
:|||||:|||||:|||||:|||||:|||||:|||||
115 spAspHisSerAsnAspLeuMetLeuHisLeuSerGlnProAlaAsp 131
268 GAGTGTGACCATCCGGAGCATCATGCTTGGCAGTGCCTACCGC 317
:|||||:|||||:|||||:|||||:|||||:|||||
132 IleThrAspGlyValLysValIleAspLeuProThrGluGluProLysVa 148
318 GGGAACTCTTGCTCGTTCTTGCTGGGCT..... 348
:|||||:|||||:|||||:|||||:|||||:|||||
148 IGLysThrCysLeuAlaSerGlyTrpGlySerThrLysProLeuIleT 165
349CTGCTGGCGAAC 360
165 rpluPheProAspAspLeuGlnCysValAsnIleHisLeuLeuSerAsn 181
361 GATGCTGTGATGCCATCCAGTCCGACACTGTG 393
:::|||||:|||||:|||||:|||||:|||||:|||||
182 GluLysCysIleLysAlaTyrLysGluLysVal 192

OM of: US-09-030-606-177 to: SPTREMBL_10:* out_format : pfs
Date: Sep 25, 1999 10:35 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09030606/runat_24091999.171617_29869/app-query.fasta.1
-DB=SPTREMBL_10 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-177
Query length: 1119
Database: SPTREMBL_10:*
Database sequences: 201082
Database length: 61543640
Search time (sec): 260.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|----------------------|------|--------|--------|---------|-----|-----------------------------------|
| sp_rodent:Q9Z0M1 | + | 469.50 | 825.82 | 1.7e-38 | 255 | Q9Z0M1 mus musculus (mouse). en |
| sp_human:Q15096 | + | 275.00 | 478.86 | 3.8e-19 | 234 | Q15096 homo sapiens (human). pr |
| sp_mammal:Q29474 | + | 271.50 | 471.64 | 8.6e-19 | 261 | Q29474 canis familiaris (dog). q1 |
| sp_human:Q15946 | + | 269.00 | 468.55 | 1.5e-18 | 223 | Q15946 homo sapiens (human). q1 |
| sp_human:Q75837 | + | 267.50 | 463.81 | 2.2e-18 | 282 | Q75837 homo sapiens (human). se |
| sp_human:Q15098 | + | 265.00 | 463.48 | 3.7e-18 | 176 | Q15098 homo sapiens (human). pr |
| sp_human:Q16272 | + | 265.00 | 462.85 | 3.7e-18 | 189 | Q16272 homo sapiens (human). ps |
| sp_rodent:Q63274 | + | 256.00 | 444.85 | 3.0e-17 | 235 | Q63274 rattus norvegicus (rat). . |
| sp_vertebrate:Q42158 | + | 254.50 | 441.74 | 4.2e-17 | 247 | Q42158 petromyzon marinus (sea |
| sp_vertebrate:Q42608 | + | 253.50 | 439.95 | 5.3e-17 | 247 | Q42608 petromyzon marinus (sea |
| sp_vertebrate:Q42160 | + | 248.50 | 431.08 | 1.7e-16 | 245 | Q42160 petromyzon marinus (sea |
| sp_rodent:Q03955 | + | 248.00 | 430.01 | 1.9e-16 | 250 | Q03955 praomys natalensis (afr |
| sp_rodent:Q63275 | + | 247.50 | 429.51 | 2.1e-16 | 239 | Q63275 rattus norvegicus (rat). . |
| sp_rodent:Q88309 | + | 245.00 | 424.27 | 3.8e-16 | 261 | Q88309 mus musculus (mouse). pr |
| sp_rodent:Q921R9 | + | 243.00 | 421.21 | 5.9e-16 | 246 | Q921R9 mus musculus (mouse). tr |
| sp_vertebrate:Q42159 | + | 242.50 | 420.39 | 6.6e-16 | 244 | Q42159 petromyzon marinus (sea |
| sp_human:Q15665 | + | 242.50 | 419.87 | 6.7e-16 | 259 | Q15665 homo sapiens (human). tr |
| sp_vertebrate:Q93265 | + | 239.00 | 413.92 | 1.5e-15 | 250 | Q93265 pleuronectes americanus. . |
| sp_rodent:Q61855 | + | 235.00 | 406.39 | 3.7e-15 | 261 | Q61855 mus musculus (mouse). tr |
| sp_rodent:Q54854 | + | 231.50 | 400.48 | 8.3e-15 | 251 | Q54854 rattus norvegicus (rat). . |
| sp_rodent:Q88301 | + | 228.00 | 394.40 | 1.8e-14 | 246 | Q88301 mus musculus (mouse). tr |
| sp_vertebrate:Q92099 | + | 227.00 | 392.75 | 2.3e-14 | 242 | Q92099 paranthothenia magellani |
| sp_vertebrate:Q93266 | + | 223.00 | 385.60 | 5.8e-14 | 242 | Q93266 pleuronectes americanus. . |
| sp_vertebrate:Q92046 | + | 223.00 | 385.35 | 5.8e-14 | 249 | Q92046 dlsostichus mawsoni. pr |
| sp_vertebrate:Q91515 | + | 222.50 | 384.89 | 6.5e-14 | 237 | Q91515 figu rubripes (japanese |
| sp_vertebrate:Q93594 | + | 212.50 | 369.92 | 6.2e-13 | 178 | Q93594 dicentrarchus labrax (eu |
| sp_human:Q99895 | + | 209.50 | 360.57 | 1.3e-12 | 268 | Q99895 homo sapiens (human). ca |
| sp_vertebrate:Q91036 | + | 207.00 | 357.87 | 2.2e-12 | 219 | Q91036 gadus morhua (atlantic c |
| sp_rodent:Q08643 | + | 207.00 | 356.23 | 3.2e-12 | 264 | Q08643 mus musculus (mouse). ly |
| sp_human:Q00765 | + | 205.50 | 353.96 | 2.3e-12 | 252 | Q00765 homo sapiens (human). ch |
| sp_vertebrate:Q13059 | + | 201.50 | 346.60 | 8.1e-12 | 258 | Q13059 trimeresurus granthicus |
| sp_mammal:Q46683 | + | 201.00 | 345.95 | 9.0e-12 | 251 | Q46683 ovis aries (sheep). mast |
| sp_rodent:Q63637 | + | 200.50 | 345.41 | 1.0e-11 | 241 | Q63637 rattus norvegicus (rat). . |
| sp_rodent:Q61388 | + | 199.50 | 343.03 | 1.3e-11 | 252 | Q61388 mus musculus (mouse). cy |
| sp_vertebrate:Q91053 | + | 199.00 | 342.22 | 1.4e-11 | 262 | Q91053 glycydus ussuriensis. ca |
| sp_vertebrate:Q92077 | + | 199.00 | 341.86 | 1.4e-11 | 266 | Q92077 gadus morhua (atlantic c |
| sp_vertebrate:Q9YGS2 | + | 195.00 | 334.91 | 3.6e-11 | 260 | Q9YGS2 glycydus halys (halys vi |
| sp_vertebrate:Q34289 | + | 194.50 | 334.19 | 4.0e-11 | 255 | Q34289 saavelinus fontinalis (h |
| sp_human:Q16018 | + | 193.00 | 333.46 | 5.0e-11 | 226 | Q16018 homo sapiens (human). ch |
| sp_vertebrate:Q9YGS1 | + | 193.00 | 332.26 | 5.6e-11 | 234 | Q9YGS1 agkistrodon acutus (dein |
| sp_rodent:Q63636 | + | 193.00 | 331.75 | 5.6e-11 | 248 | Q63636 rattus norvegicus (rat). . |
| sp_vertebrate:Q9YGS6 | + | 193.00 | 331.34 | 5.7e-11 | 260 | Q9YGS6 agkistrodon halys pallas |

sp_rodent:Q63224 + 192.00 329.96 7.1e-11 248 | Q63224 rattus norvegicus (ra
sp_human:Q14243 + 190.00 325.68 1.1e-10 269 | Q14243 homo sapiens (human).
sp_mammal:Q19023 + 189.50 325.18 1.3e-10 257 | Q19023 macaca mulatta (rhesu

seq_name: sp_rodent:Q9Z0M1

seq_documentation_block:
ID Q9Z0M1 PRELIMINARY; PRT; 255 AA.
AC Q9Z0M1;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS-WEBSTER;
RA Simmer J.;
RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF019979; AAC98894.1; .
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.
SQ SEQUENCE 255 AA; 27488 MW; 6E711616 CRC32;

alignment_scores:
Quality: 469.50 Length: 139
Ratio: 4.083 Gaps: 1
Percent Similarity: 82.734 Percent Identity: 63.309

alignment_block:
US-09-030-606-177 x Q9Z0M1 ..

Align seg 1/1 to: Q9Z0M1 from: 1 to: 255

4 CACTGCGAGCCCTGGCAGCGCGGCGCATGTCATGGAACGAATGTTCTG 53
|||||
41 HisserGlnProTrrpGlnAlaAlaLeuPheSerGlnAlaAlaHisCysl 57
54 CTCGGCGCTCTGTCATCCGCGAGTGGTGTGTCAGCCGACACTGTT 103
|||||
57 sserGlyValLeuValHisProGlnTrrpValLeuSerAlaAlaHisCysl 74
104 TCCAGACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
:::|||||
74 eugInGlnSerTrrpGlnValGlyLeuGlyLeuHisAsnLeuLysGlySer 90
154 CAAGAGCCAGGAGCCAGATGTTGAGGCCAGCCTCTCCGATGACGCC 203
|||||
91 GlnGlnProGlnSerArgMetLeuGlnAlaHisLeuSerIleGlnHisPr 107
204 AGAGTACACAGACCCCTGCTGCTGCTAAGCAGCCTCATGCTCAAGTTGG 253
|:::|
107 oAsnPheAsnAspProSerPheAlaAsnAspLeuMetLeuIleLysLeuA 124
254 ACGAATCCGTCGAGTCTGACACCATCCGAGCATCAGCATGCTTCTCG 303
:::|
124 snGlnSerValIleGlnSerAsnTrpIleArgSerIleProValAlaThr 140
304 CAGTGCCCTACCGCGGAGCACTGCTGCTGCTTCTGCTGGGCTGCTGCT 353
|||||
141 GlnCysProThrProGlnAspTrpCysLeuValSerGlyTrrpGlyGlnLe 157
354 GCGGACGATGCTGTGATGTCATCCAGTCCAGACTGTGAGGAGGCTGGG 403
| |
157 ulysAsnGlyLysLeuProSerLeu.....LeuG 167
404 AGTGTGAGAACTTTCC 420
::| |

167 IncysValAsnLeuser 172

seq_name: sp_human:Q15096

seq_documentation_block:

ID Q15096 PRELIMINARY; PRT; 234 AA.
AC Q15096;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)
DE PROSTATE SPECIFIC ANTIGEN (PSA) PRECURSOR (FRAGMENT).
GN APS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88326297.
RA RIEGMAN P.H., KLAASSEN P., DER KORPUT J.A., ROMIJN J.C., TRAPMAN J.;
RT "Molecular cloning and characterization of novel prostate antigen
RT cDNA's."
RL Biochem. Biophys. Res. Commun. 155:181-188(1988).
DR EMBL; M21896; AAA5996.1; -.
DR PFM; PF00089; trypsin; 1.
FT NON_TER 1 1
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 21 234 POTENTIAL.
SQ SEQUENCE 234 AA; 25803 MW; 1B244D8C CRC32;

alignment_scores:

Quality: 275.00 Length: 179
Ratio: 2.670 Gaps: 4
Percent Similarity: 57.542 Percent Identity: 34.078

alignment_block:

US-09-030-606-177 x Q15096 ..

Align seg 1/1 to: Q15096 from: 1 to: 234

4 CACTCGAGCCCTGGCAGCGGCACACTGTCATGGAAGAATGTCTG 53
|||||
30 HissSerGlnProTrrpGlnValLeuValAlaSerArgGlyArgAlaValCys 46
54 CTGGGGCGTCTGTGTCATCCGACAGTGGGTGCTTCACGCCGACACTGT 103
|
46 sglGlyValLeuValHisProGlnTrrpValLeuThrAlaAlaHisCys 63
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGGCCGAC 153
|
63 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 79
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACC 203
|
80 GluAspThrGly..GlnValPheGlnValSerHisSerPheProHisPr 95
204 AGAGTACAACAGACCCTTGCTCGCTAAC..... 231
|
95 oleuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAsp 112
232GACCTCATGCTCATCAAGTTGAGCAATCCGTGTCGAG 270
|
112 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 128
271 TCTGACACCATCCGAGACATCAGCATGTGCTTCGACAGTCCCTACCGCGG 320
|
129 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG 145
321 GAAGTCTTGCTGCTTCTGCTGGGT..... 348
|
145 YThrThrCysTyrAlaSerGlyTrrpGlySerIleGluProGluGluPhe 162
348CTGCTGGCGAAGCAT 363

162 eutHrProLysLysLeuGlnCysValAspLeuHisValIleSerAsnAsp 178

seq_name: sp_mammal:Q29474

seq_documentation_block:

ID Q29474 PRELIMINARY; PRT; 261 AA.
AC Q29474;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)
DE KALLIKREIN PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (GLANDULAR KALLIKREIN) (KININOGENIN).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MONGREL;
RX MEDLINE; 94250683.
RA GAUTHIER E.R., DUMAS C., CHAPDELAIN P., TREMBLAY R.R., DUBE J.Y.;
RT "Characterization of canine pancreas kallikrein cDNA."
RL Biochim. Biophys. Acta 1218:102-104(1994).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES.
CC HIGHLY SELECTIVE ACTION TO RELEASE KALLIDIN
CC (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC THE RAT ENZYME IS UNUSUAL IN LIBERATING BRADYKININ DIRECTLY FROM
CC AUTOLOGOUS KININOGENS BY CLEAVAGE AT TWO ARG-|-XAA BONDS.
DR EMBL; X75479; CAA53210.1; -.
DR PFM; PF00089; trypsin; 1.
KW signal; hydrolase.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 261 KALLIKREIN.
SQ SEQUENCE 261 AA; 28913 MW; 65B49AEF CRC32;

alignment_scores:

Quality: 271.50 Length: 230
Ratio: 2.207 Gaps: 8
Percent Similarity: 53.478 Percent Identity: 31.739

alignment_block:

US-09-030-606-177 x Q29474 ..

Align seg 1/1 to: Q29474 from: 1 to: 261

4 CACTCGAGCCCTGGCAGCGGCACACTGTCATGGAAGAATGTCTG 53
|
34 AsnSerGlnProTrrpGlnAlaAlaLeuTyrHisTyrSerLysPheGlnCys 50
54 CTGGGGCGTCTGTGTCATCCGACAGTGGGTGCTTCACGCCGACACTGT 103
|
50 sglGlyValLeuValHisProGluTrrpValValThrAlaAlaHisCys 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
|
67 leAsnAspAsnTyrGlnLeuTrrpLeuGlyArgTyrAsnLeu..PheGlu 82
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACC 203
|
83 HisGluAspThrAlaGlnPheValGlnValArgGluSerPheProHisPr 99
204 AGAGTACAACAGACCCTTGCTCGCTAAC..... 231


```
||||:|||||
99 ogLpHeAsnLeuSerLeuLeuLysAsnHisThrArgLeuProGluGluA 116
232 .....GACCTCATGCTCATCAAGTTGGAGCAATCCGTGCCGAG 270
116 sPTySerHisAspIleMetLeuLeuArgLeuAlaGluProAlaGlnIle 132
271 TCTGACACCATCCGGAGCATCAGCATTGGCTTGCAGTGCCCTACCGCGG 320
133 ThrAspAlaValArgValLeuAspLeuProThrGlnGluProGlnValG1 149
321 GAACCTCTTGCCTGTTTCTGGCTGGGT..... 348
149 ySerThrCysTyAlaSerGlyTrpGlySerIleGluProAspLysPhe1 166
349 .....CTGCTGGCGAAGCAT 363
166 lETyRProAspAspLeuGlnCysValAspLeuGlnLeuLeuSerAsnAsp 182
364 GCTGTGATGTCATCCAGTCCAGACTGTG..... 393
183 IleCysAlaAsnAlaHisSerGlnLysValThrGluPheMetLeuCysAl 199
394 .....GGAGGCTGGGAG..TGTGAGAAGCTTTCCCA..... 422
199 agLyHisLeuGlnGlyLysAspThrCysValGlyAspSerGlyGlyP 216
423 .....ACCCTGGCAGGGTTGTACCATTTGCGCAACTTC 455
216 roLeuIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlyHisVal 232
456 CAGTGC.....AAGAGCTCCTGCTGCA 478
233 ProCysGlySerProAsnMetProAlaValTyThrLysValIleSerH1 249
479 TCCTCACTGGGTGCTCACTACTGCTCACTGCATCACCC 516
249 sLeuGluTrpIleLysGluThr.MetThrAlaAsnPro 261
```

seq_name: sp_human:Q15946

```
seq_documentation_block:
ID Q15946 PRELIMINARY; PRT; 223 AA.
AC Q15946;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE GLANDULAR KALLIKREIN-1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92324494.
RA RIEGMAN P.H., VLIETSTRA R.J., VAN DER KORPUT H.A., ROMIJN J.C.,
RA TRAPMAN J.;
RT "Identification and androgen-regulated expression of two major human
RT glandular kallikrein-1 (hGK-1) mRNA species.";
RL Mol. Cell. Endocrinol. 76:181-190(1991).
DR EMBL; S39329; AAD13817.1; -.
DR PFW; PF00089; trypsin; 1.
SQ SEQUENCE 223 AA; 24674 MW; E0321503 CRC32;
```

```
alignment_scores:
Quality: 269.00 Length: 183
Ratio: 2.637 Gaps: 4
Percent Similarity: 55.738 Percent Identity: 33.880
```

alignment_block:

US-09-030-606-177 x Q15946 ..

Align seg 1/1 to: Q15946 from: 1 to: 223

```
4 CACTCGACGCCCTGGCAGCGCGCACCTGTCATGGAAACGAATGTTCTG 53
|||||
34 HisSerGlnProTrpGlnValAlaValTySerHisGlyTrpAlaHisCys 50
54 CTCGGCGCTGCTGGTCATCCGAGTGGTCTGTCCAGCCGACACTGTT 103
|||||
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 67
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
|||||
67 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 83
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCACCTCTCCGTACCGCACCC 203
|||||
84 GluAspThrGly..GlnArgValProValSerHisSerPheProHisPr 99
204 AGAGTACACACAGACCTTGCTC..... 225
99 oLeuTyRAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 116
226 ....GCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCCGAG 270
|||||
116 sPserSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 132
271 TCTGACACCATCCGGAGCATCAGCATTGGCTTGCAGTGCCCTACCGCGG 320
|||||
133 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuG1 149
321 GAACCTCTTGCCTGTTTCTGGCTGGGT..... 348
|||||
149 yThrThrCysTyAlaSerGlyTrpGlySerIleGluProGluGluPhe1 166
349 .....CTGCTGGCGAAGCAT 363
166 euArgProArgSerLeuGlnCysValSerLeuHisLeuSerAsnAsp 182
364 GCTGTGATGTCATCCAGTCCAGACTGTG..... 393
183 MetCysAlaArgAlaTySerGluLysValThrGluPheMetLeuCysAl 199
394 .....GGAGGCTGGAGTGTGAGAAGCTTTCCCAACCTGG 429
199 agLyLeuTrpThrGlyGlyLysAspThrCysGlyValSerHisProTyR 215
```

seq_name: sp_human:075837

```
seq_documentation_block:
ID 075837 PRELIMINARY; PRT; 282 AA.
AC 075837;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE SERINE PROTEASE (TLSP).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98438738.
RA YOSHIDA S., TANIGUCHI M., SUEMOTO T., OKA T., HE X.P., SHIOSAKA S.;
RT "cDNA cloning and expression of a novel serine protease, TLSP.";
RL Biochim. Biophys. Acta 1399:225-228(1998).
DR EMBL; AB012917; BAA33404.1; -.
DR PFW; PF00089; trypsin; 1.
KW Protease; Serine protease.
FT CHAIN 54 282 TLSP.
SQ SEQUENCE 282 AA; 31059 MW; E97E294C CRC32;
```

```
alignment_scores:
Quality: 267.50 Length: 119
Ratio: 3.075 Gaps: 2
```


Percent Similarity: 73.109 Percent Identity: 43.697

alignment_block:

US-09-030-606-177 x Q15098 ..

Align seg 1/1 to: 075837 from: 1 to: 282

```
4 CACTCGCAGCCCTGGCAGCGGCACACTGGTCATGGAAGCAATTGTCTG 53
|||||
63 HisSerGlnProTyrGlnAlaAlaLeuPheGlnLysThrArgLeuLeu 79
54 CTCGGGCGCTCCTGTGTCATCCGACAGTGGTGTCTGACCGCACACTGT 103
|||||
79 sGlyAlaThrLeuIleAlaProArgTyrLeuLeuThrAlaAlaHisCysL 96
104 TCCAGAACTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGCGCAC 153
|||||
96 euLysProArgTyrIleValHisLeuGlnHisAsnLeuGln..Lys 111
154 CAAGAGCCAGGAGCCAGATGTGTGAGCCAGCCCTCTCCGTACGGCACC 203
|||||
112 GluGlnGlyCysGlnThrArgThrAlaThrGlnSerPheProHisPr 128
204 AGAGTACAACAGACCCCTG.....CTCCGTAACGACCTCATGC 241
|||||
128 OGlyPheAsnAsnSerLeuProAsnLysAspHisArgAsnAspIleMet 145
242 TCATCAAGTTGGACGAATCCGTCTCCGAGTCTGACACCATCCGAGCATC 291
|||||
145 euValLysMetAlaSerProValSerIleThrTrpAlaValArgProLeu 161
292 AGCATGTCTTCGACGTCCCTACCGCGGGAACCTTGCCCTGTTCTGG 341
|||||
162 ThrLeuSerSerArgCysValThrAlaGlyThrSerCysLeuIleSerG 178
342 CTGGGGT 348
|||||
178 YTrpGly 180
```

seq_name: sp_human:Q15098

seq_documentation_block:

```
ID Q15098 PRELIMINARY; PRT; 176 AA.
AC Q15098;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROSTATE SPECIFIC ANTIGEN (PSA) PRECURSOR (FRAGMENT).
GN APS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88326297.
RA RIEGMAN P.H., KLAASSEN P., DER KORPUT J.A., ROMIJN J.C., TRAPMAN J.;
RT "Molecular cloning and characterization of novel prostate antigen
RT cDNA's."
RL Biochem. Biophys. Res. Commun. 155:181-188(1988).
DR EMBL; M21897; AAAS9997.1; -.
DR PFAM; PF00089; trypsin; 1.
FT NON_TER 1
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 21 176 POTENTIAL.
SQ SEQUENCE 176 AA; 19158 MW; 949BC6C3 CRC32;
```

alignment_scores:

Quality: 265.00 Length: 126
Ratio: 3.011 Gaps: 2
Percent Similarity: 69.841 Percent Identity: 42.063

alignment_block:

US-09-030-606-177 x Q15098 ..

Align seg 1/1 to: Q15098 from: 1 to: 176

```
4 CACTCGCAGCCCTGGCAGCGGCACACTGGTCATGGAAGCAATTGTCTG 53
|||||
30 HisSerGlnProTyrGlnValLeuValAlaSerArgGlyArgAlaValCys 46
54 CTCGGGCGCTCCTGTGTCATCCGACAGTGGGTGCTGTACGCCGACACTGT 103
|||||
46 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 63
104 TCCAGAACTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGCGCAC 153
|||||
63 leArgAsnLysSerValIleLeuLeuGlnArgHisSerLeuPheHisPro 79
154 CAAGAGCCAGGAGCCAGATGTGTGAGCCAGCCCTCTCCGTACGGCACC 203
|||||
80 GluAspThrGly..GlnValPheGlnValSerHisSerPheProHisPr 95
204 AGAGTACAACAGACCCCTGCTCGCTAAC..... 231
|||||
95 OLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAsp 112
232 .....GACCTCATGCTCATCAAGTTGACGAATCCGTGTCGAG 270
|||||
112 spSerSerHisAspLeuMetLeuLeuArgLeuSerGlnProAlaGlnLeu 128
271 TCTGACACCATCCGAGCATCAGCATGCTTCGCAAGTGCCTTACCGCGG 320
|||||
129 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG 145
321 GAACCTTGCCCTGTTCTTGCGTGGGT 348
|||||
145 YThrThrCysTyrAlaSerGlyTrpGly 154
```

seq_name: sp_human:Q16272

seq_documentation_block:

```
ID Q16272 PRELIMINARY; PRT; 189 AA.
AC Q16272;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PSA PROTEIN (FRAGMENT).
GN PSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95079406.
RA MONNE M., CROCE C.M., YU H., DIAMANDIS E.P.;
RT "Molecular characterization of prostate-specific antigen messenger
RT RNA expressed in breast tumors."
RL Cancer Res. 54:6344-6347(1994).
DR EMBL; S75755; AAD14185.1; -.
DR PFAM; PF00089; trypsin; 1.
FT NON_TER 1
FT SEQUENCE 189 AA; 21039 MW; DB4C6482 CRC32;
```

alignment_scores:

Quality: 265.00 Length: 126
Ratio: 3.011 Gaps: 2
Percent Similarity: 69.841 Percent Identity: 42.063

alignment_block:

US-09-030-606-177 x Q16272 ..
Align seg 1/1 to: Q16272 from: 1 to: 189

4 CACTCGCAGCCCTGGCAGCGGCACACTGGTCATGGAAGCAATTGTCTG 53


```
|||||
47 HisSerGlnProTrrpGlnValLeuValAlaSerArgGlyArgAlaValCys 63
54 CTCGGCGCTCCTGGTGCATCCGACAGTGGGTGCTGCAGCCCGCACACTGTT 103
   |||||
63 sGlyGlyValLeuValHisProGlnTrrpValLeuThrAlaAlaHisCysI 80
104 TCCAGAACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCCGAC 153
   |||||
80 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 96
154 CAAGAGCCAGGAGCCAGATGGTGGAGCCAGCCTCTCCGTAGCGCACCC 203
   |||||
97 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 112
204 AGAGTACAACAGACCCTTGCTCGCTAAC..... 231
   |||||
112 OleuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 129
232 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAG 270
   |||||
129 sPserSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 145
271 TCTGACACCATCCGGAGCATCAGCATGCTTGCAGTGCCTACCGCGGG 320
   |||||
146 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG 162
321 GAACCTTGCTCGTCTTCTGCTGGGT 348
   |||||
162 yThrThrCysTyrAlaSerGlyTrrpGly 171
seq_name: sp_rodent:Q63274
```

```
seq_documentation_block:
ID Q63274 PRELIMINARY; PRT; 235 AA.
AC Q63274;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
RA ZIMTZ C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33839; AAA58781.1; -.
DR PFM; PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 235 AA; 26226 MW; EGDE5AD8 CRC32;
```

```
alignment_scores:
Quality: 256.00 Length: 126
Ratio: 2.943 Gaps: 3
Percent Similarity: 69.048 Percent Identity: 41.270
```

```
alignment_block:
US-09-030-606-177 x Q63274 ..
```

```
Align seg 1/1 to: Q63274 from: 1 to: 235
4 CACTCGAGCCCTGGCAGGCGGCACTGTCATGAAACGAATTGTTCTG 53
   |||||
10 AsnSerGlnProTrrpGlnValAlaValIle.....AsnArgTyrLeuCy 24
54 CTCGGCGCTCCTGGTGCATCCGACAGTGGGTGCTGCAGCCCGCACACTGTT 103
   |||||
24 sGlyGlyValLeuLeuAspProSerTrrpValIleThrAlaAlaHisCysT 41
104 TCCAGAACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCCGAC 153
```

```
..... ||| :: ||||| ||||| |||
41 yrrSerHisTrrTrrHisValLeuLeuGlyArgHisAsnLeuPheGluAsp 57
154 CAAGAGCCAGGAGCCAGATGTTGGAGGCCAGCCTTCCTCCGTACGGCACCC 203
   |||||
58 ..GluProPheAlaGlnTyrArgPheValSerGlnSerPheProHisPr 73
204 AGAGTACAACAGACCCTTGCTC..... 225
   |||||
73 oAspTyrAsnProPheLeuMetArgAsnHisThrArgGlnThrGlyTyrA 90
226 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAG 270
   |||||
90 sPtyrSerAsnAspLeuMetLeuLeuHisLeuSerGluProAlaAspIle 106
271 TCTGACACCATCCGGAGCATCAGCATGCTTCCGACGTGCCCTACCGCGG 320
   |||||
107 ThrAspGlyValLysValIleAspLeuProThrGlnGluProLysValG 123
321 GAACCTTGCTCGTCTTCTGCTGGGT 348
   |||||
123 ySerThrCysLeuValSerGlyTrrpGly 132
seq_name: sp_vertebrate:042158
```

```
seq_documentation_block:
ID 042158 PRELIMINARY; PRT; 247 AA.
AC 042158;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPSINOGEN A2 PRECURSOR.
GN TRYP2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011898; AAB69654.1; -.
DR PFM; PF00089; trypsin; 1.
KW SIGNAL.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 TRYPSIN A2.
SQ SEQUENCE 247 AA; 26309 MW; D049AA4C CRC32;
```

```
alignment_scores:
Quality: 254.50 Length: 143
Ratio: 2.447 Gaps: 4
Percent Similarity: 72.727 Percent Identity: 34.965
```

```
alignment_block:
US-09-030-606-177 x 042158 ..
```

```
Align seg 1/1 to: 042158 from: 1 to: 247
1 GCGCACTCGAGCCCTGGCAGGCGGCACTGTCATGAAACGAATTGTT 50
   |||||
32 AlaHisSerGlnProTrrpGlnValSerLeuAsnIleGlyTrrHis...Ph 47
51 CTGCTGGCGCTCCTGGTGCATCCGACAGTGGGTGCTGCAGCCCGCACACT 100
   |||||
47 ecysGlyGlySerLeuLeuLeuAsnSerGlnTrrpValValSerAlaAlaHisC 64
101 GTTCCAGAACTCC.....TACACCATCGGCTGGGCTGCACAGTCTT 144
   |||||
64 ystYrGlnThrAlaSerArgIleSerValArgIleGlyGluHisAsnIle 80
145 GAGCGCGACCAAGAGCCAGGAGCCAGATGTTGGAGGCCAGCCTTCCT 194
   |||||
81 pheValAsn...GluGlyThrGlnGlnIleGlnAlaSerLysAlaIle 96
```



```
195 ACGGCACCCAGAGTACACAGACCCTTGCTCGCTACGACCTCATGCTCA 244
    :::::::::::::::::::::
96 eGlnHisProGlnTyraSerTrpThrIleaspAsnAspIleMetLeuI 113
245 TCAGTTGGACGAATCCGTCCTCCAGTCTGACACCATCCGAGCATCAGC 294
    |||||||:
113 IeIysLeuSerSerProAlaThrLeuAsnGlnTyraIaGlnAlaIleAla 129
295 ATTGCTTCGCAGTGCCTTACCGCGGGAACCTCTGCTTCTTGCGCTG 344
    :::: |||:
130 LeuProSerSerCysValAsnThrGlyValMetCysThrIleSerGlyTr 146
345 GGGT.....CTGCTGGCGAAGCATGCTGTGATGCCATCC 379
    |||:
146 pGlyGluThrGlnThrSerIleGlySerProAspValLeuMetCysValG 163
380 AGTCCACGACTGTGGAGGCTGGAGTGT 408
    |||:
163 IAlaProValLeuSerAspThrSerCys 172
```

seq_name: sp_vertebrate:042608

seq_documentation_block:

```
ID 042608 PRELIMINARY; PRT; 247 AA.
AC 042608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPsinGEN A3 PRECURSOR.
GN TRYP3.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=ANTERIOR INTESTINE;
RA ROACH J.C.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011899; AAB69655.1; -.
DR EMBL; AF011352; AAB65411.1; -.
DR PFAM; PF00089; trypsin; 1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 TRYPsin A3.
SQ SEQUENCE 247 AA; 26295 MW; 6D71AC2E CRC32;
```

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 253.50 | Length: | 143 |
| Ratio: | 2.438 | Gaps: | 4 |
| Percent Similarity: | 72.727 | Percent Identity: | 34.965 |

alignment_block:

US-09-030-606-177 x 042608 ..

Align seg 1/1 to: 042608 from: 1 to: 247

```
1 GCGCACTCGCAGCCCTGGCAGCGGCGCACTGTCATGGAAGAATGTT 50
    |||||||:
32 AlaHisSerGlnProTrpGlnValSerLeuAsnIleGlyTyrHis..Ph 47
51 CTGCTCGGGCGTCTGTGTCATCCGAGTGGGTGCTGTACGCCGACACT 100
    |||:
47 eCysGlyGlySerLeuIleAsnSerGlnTrpValValSerAlaAlaHisC 64
101 GTTTCAGAACTCC.....TACACCATCGGGCTGGCCTGCACAGTCTT 144
    |||:
64 yGtyrGlnThrAlaSerArgIleSerValArgIleGlyGluHisAsnIle 80
```

```
145 GAGCCGACCAAGACCCAGGAGCCAGATGCTGAGGCCAGCTCTCCGT 194
    ::::: |||:
81 pheValasn..GluGlyThrGluGlnGlnIleGlnAlaSerIysAlaI 96
195 ACGGCACCCAGAGTACCAACAGACCCTTGCTCGCTAACGACCTCATGCTCA 244
    :::::::::::::::::::::
96 eGlnHisProGlnTyraSerTrpThrIleaspAsnAspIleMetLeuI 113
245 TCAGTTGGACGAATCCGTCCTCCAGTCTGACACCATCCGAGCATCAGC 294
    |||||||:
113 IeIysLeuSerSerProAlaThrLeuAsnGlnTyraIaGlnAlaIleAla 129
295 ATTGCTTCGCAGTGCCTTACCGCGGGAACCTCTGCTTCTTGCGCTG 344
    :::: |||:
130 LeuProSerSerCysValAsnThrGlyValMetCysThrIleSerGlyTr 146
345 GGGT.....CTGCTGGCGAAGCATGCTGTGATGCCATCC 379
    |||:
146 pGlyGluThrGlnThrSerValGlySerProAspValLeuMetCysValG 163
380 AGTCCACGACTGTGGAGGCTGGAGTGT 408
    |||:
163 IAlaProValLeuSerAspThrSerCys 172
```

seq_name: sp_vertebrate:042160

seq_documentation_block:

```
ID 042160 PRELIMINARY; PRT; 245 AA.
AC 042160;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPsinGEN B2 PRECURSOR (FRAGMENT).
GN TRYPB2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011901; AAB69657.1; -.
DR PFAM; PF00089; trypsin; 1.
KW Signal.
FT NON_TER 1 1 POTENTIAL.
FT SIGNAL <1 13 TRYPsin B2.
FT CHAIN 14 245
SQ SEQUENCE 245 AA; 26001 MW; 071872F0 CRC32;
```

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 248.50 | Length: | 143 |
| Ratio: | 2.510 | Gaps: | 4 |
| Percent Similarity: | 69.231 | Percent Identity: | 34.965 |

alignment_block:

US-09-030-606-177 x 042160 ..

Align seg 1/1 to: 042160 from: 1 to: 245

```
1 GCGCACTCGCAGCCCTGGCAGCGGCGCACTGTCATGGAAGAATGTT 50
    |||||||:
30 AlaHisSerGlnProTrpGlnValSerLeuAsnIleGlyTyrHis..Ph 45
51 CTGCTCGGGCGTCTGTGTCATCCGAGTGGGTGCTGTACGCCGACACT 100
    |||:
45 eCysGlyGlySerLeuIleAsnSerGlnTrpValValSerAlaAlaHisC 62
101 GTTTCAGAACTCC.....TACACCATCGGGCTGGCCTGCACAGTCTT 144
    |||:
62 yTyrGlnThrAlaSerArgIleSerValArgIleGlyGluHisAsnIle 78
```



```

145 GAGGCCGACCAAGAGCCAGGCCAGATGTTGGAGGCCAGCTCTCCGT 194
      ::      |||      ::|||      ::|||      ::|||      ::|||
79  pheVal..ThrGluGlyThrGluGlnArgIleGlnAlaSerLysAlaI 94

195 ACGGCACCCAGAGTACACACAGACCTTGCTCGCTAACGACCTCATGCTCA 244
      ::|||      ::|||      ::|||      ::|||      ::|||
94  earGHisProGlnTyrAsnSerAlaThrIleAspAsnAspIleMetLeu 111

245 TCAAGTTGGAGCAATCCGTGTCCGAGTCTGACACACCATCCGGAGCATCAGC 294
      |||||      ::|||      ::|||      ::|||      ::|||
111  IeLysLeuSerSerProAlaThrLeuAsnGlnTyrAlaGlnAlaIlePro 127

295 ATTGCTTCGCAGTGCCTTACCAGCGGGGGAACCTCTGCCTCGTTCTGGCTG 344
      ::|||      ::|||      ::|||      ::|||      ::|||
128  LeuProSerSerCysValGlyThrClyValMetCysThrIleSerGlyTr 144

345 GGGT.....CTGCTGGCCGAACGATGCTGTGATTGCCATCC 379
      ||||      ::|||      ::|||      ::|||      ::|||
144  pGlyGluThrGlnThrSerValGlySerProAspValLeuMetCysValG 161

380 AGTCCCAGACTGTGGGAGGCTGGGAGTGT 408
      ||:::      ::|||      ::|||      ::|||
161  InAlaProValLeuSerAspThrSerCys 170

```

seq_name: sp_rodent:Q03955

```

seq_documentation_block:
ID      003955      PRELIMINARY;      PRT;      250 AA.
AC      003955;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT      01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE      GLANDULAR KALLIKREIN (FRAGMENT).
OS      Praomys natalensis (African soft-furred rat) (Mastomys natalensis).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mastomys.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-SALIVARY GLAND;
RX      MEDLINE; 94226702.
RA      BELL R.A., FAHNESTOCK M.;
RT      "Characterization of kallikrein cDNAs from the African rodent
RT      Mastomys.";
RL      DNA Cell Biol. 13:293-300(1994).
CC      -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC      IN KININOGEN TO RELEASE LYS-BRADYKININ.
DR      EMBL; X17351; CAA35231.1; -.
DR      PFM; PF00089; trypsin; 1.
KW      Hydrolase; Serine protease; Kininogenase; Glycoprotein; Saliva;
KW      Submandibular gland; Multigene family; Zymogen; Signal.
FT      NON_TER      1      1
SQ      SEQUENCE      250 AA;      27385 MW;      BD8435E0 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 248.00 | Length: 126 |
| Ratio: | 2.787 | Gaps: 2 |
| Percent Similarity: | 70.635 | Percent Identity: 37.302 |

alignment_block:

US-09-030-606-177 x Q03955

Align seg 1/1 to: 003955 from: 1 to: 250

4 CACTCCGACCCCTGGCAGCGCGGCACTGGTCATGGAAACGAATTGTTCTG 53
:::|||||:::|||||::: ::::: |||
31 AsnSerGlnProTrpHisValAlaValTyrTyrPheargThrTyrIleCy 47
54 CTGGGGCGTCTGTGTCATCCGCGAGTGGGTGCTGTACGGCGCACACTGTT 103
|:::|||||::: ||:::|||||:::|||||::: |||
47 sGlyGlyValIeuLeuAspProAsnTrpValIeuThrAlaAlaHisCysT 64
104 TCCAGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153

```

64 yrrhrThrTyrTyrGlnValTrpLeuGlyLysAsnAsnLeuLys...Glu 79
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCCTCTCCGTAAGGCCACC 203
      |||||:::||||: |||::||| :::: |||||
80 TyrGluProSerAlaGlnPheArgGluValSerLysAlaIleProHisPr 96
204 AGAGTACAC.....AGACCCT 220
      |:::||
96 oglyPheAsnPheserAsnmetAspGlnThrThrGlnThrLysLysA 113
221 TGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAG 270
      ::||| |||||:::||||: ||| ::::
113 splyrSerAsnAspLeuMetLeuLeuArgLeuSerGluProAlaGluIle 129
271 TCTGACACCATCCGGAGCATCAGCATTGCTTCGACAGTGCCCTACGGGG 320
      ::|||:::||||: |||::: :::: |||||:::|
130 ThrAspValValLysProIleAspLeuProThrGluGluProThrValGl 146
321 GAACCTTGCCCTCGTTTCTGGCTGGGGT 348
      |::: |||||:::||||| |||||
146 yserArgCysLeuAlaSerGlyTyrpGly 155

```

seq_name: sp_rodent:Q63275

```

seq_documentation_block:
ID      063275      PRELIMINARY;      PRT;      239 AA.
AC      Q63275;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT      01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE      KALLIKREIN (FRAGMENT).
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathl; Muridae; Murinae; Rattus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RA      ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL      Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR      EMBL; L33840; AAA58782.1; -.
DR      PFAM; PF00089; trypsin; 1.
FT      NON_TER      1
SQ      SEQUENCE      239 AA;      26382 MW;      0609E3E6 CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 247.50 | Length: 160 |
| Ratio: | 2.500 | Gaps: 6 |
| Percent Similarity: | 61.875 | Percent Identity: 35.000 |

alignment_block:

US-09-030-606-177 x Q63275

Align seg 1/1 to: Q63275 from: 1 to: 239

```

4  CACTGCGACGCCCTGGCAGGGCGGCACTGGTCATGTGAAACGCAATTGTTCTG  53
   ::::::::::::::::::::::::::::  |||||  ::||
18  AsnSerGlnProTyrGlnValAlaValIle....AsnGlnAspLeuCy  32
   ::::::::::::::::::::::::::::  |||||  ::||
54  CTCGGGCGTCCCTGGTGCAATCCGCACTGGGTGCTGTACGCCGCACTGTT  103
   ||::::::::::::::::::  ||::::::::::|||::::|
32  sGlyGlyValIleuIleAspProSerTyrValIleThrAlaAlaHisCysT  49
   ::::::::::::::::::::  ::  |||||  ::::::|
104  TCCAGAACTCCTACACCAATCGGGCTGGGCCCTGCACAGTCTTGAGGCCGAC  153
   ::::::::::::::|  ::  |||||  ::::::|
49  yTserAspAsnTyrHisValLeuGlyGlnAsnAsnLeuSerGluAsp  65
   ::::::::::::::|  ::  |||||  ::::::|
154  CAAGAGCCAGGGAGGCCAGATGTGGAGGGCCAGCCTTCCGTAGGCCACCC  203
   ::  ::::::|  ||  ||  |||||
66  ValGln....HisArgLeuVal....SerGlnSerPheArgHisPr  78
   ::::::::::::::::::::  ::::::::::::::::::::
204  AGAGTACAACAGACCCCTTGCTC.....  225

```



```

||||| ::|||::|||
78 oAspTyr...LysProPheLeuMetaLrgAsnHisThrArgLysProLysA 94
226 ....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
94 sPtyrSerAsnAspLeuMetLeuHisLeuSerGluProAlaAspIle 110
271 TCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGCCTACCGCGGG 320
111 ThrAspGlyValLysValIleAspLeuProThrLysGluProLysValGl 127
321 GAACCTCTGCCTCGTTTCTGCGTGGGT..... 348
127 ySerThrCysLeuValSerGlyTyrGlySerThrAsnProSerGluTyrG 144
349 .....CTGCTGGCGAAGCAT 363
144 IupheProAspAspLeuGlnCysValAsnIleHisLeuSerAsnGlu 160
364 GCTGTGATTGCCATCCAGTCCCGACTGTG 393
161 LysCysIleLysAlaTyrLysGluLysVal 170
seq_name: sp_rodent:O88309
```

```

seq_documentation_block:
ID 088309 PRELIMINARY; PRT; 261 AA.
AC 088309;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PRORENIN-CONVERTING ENZYME (MK13B) PRECURSOR (MK13B).
GN MK1K-13B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98351995.
RA HOSOI K., TADA J., TSUMURA K., KANAMORI N., YAMANAKA N.;
RT "Expression of an allozyme of prorenin-converting enzyme in the
RT submandibular gland of DBA/2N mice.";
RL J. Biochem. 124:368-376(1998).
DR EMBL; AB016032; BAA31686.1; -.
DR PFW; PF00089; trypsin; 1.
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
SQ SEQUENCE 261 AA; 28692 MW; 8DB1814F CRC32;
```

```

alignment_scores:
Quality: 245.00 Length: 126
Ratio: 2.784 Gaps: 2
Percent Similarity: 69.841 Percent Identity: 38.889
```

alignment_block:

US-09-030-606-177 x O88309 ..

Align seg 1/1 to: O88309 from: 1 to: 261

```

4 CACTCGACGCCCTGGCAGGCGGCACTGGTCATGAAAAAGCAATGTCTTG 53
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
34 AsnSerGlnProTyrGlnValAlaValTyrTyrGlnLysGluHisIleCys 50
54 CTCGGGCGTCTGTGTCATCCGACAGTGGTGTCTGACGCCGACACTGTT 103
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
50 sGlyGlyValLeuLeuAspArgAsnTyrValLeuThrAlaAlaHisCyst 67
104 TCCAGAACTCCTACACCATCGGGCTGGCTGCACAGCTCTTGAGCCGAC 153
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
67 yRValAspGlnTyrGlnValTyrPLeuGlyLysAsnLysLeu...PheGln 82
** 154-CATGAGCCAGGAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGCAACC 203
```

```

:::|||||:::|||||:::|||||:::|||||:::|||||
83 GluGluProSerAlaGlnHisArgLeuValSerLysSerPheProHisPr 99
204 AGAGTACAACAGACCCTTGCTC..... 225
99 oGlyTyrAsnMetSerLeuMetLeuGlnThrIleProProGlyAla 116
226 ....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
116 sPheSerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIle 132
271 TCTGACACCATCCGAGCATCAGCATTGCTTCGAGTGCCTACCGCGGG 320
133 ThrAspValValLysProIleAlaLeuProThrLysGluProLysProGl 149
321 GAACCTCTGCCTCGTTTCTGCGTGGGT 348
149 ySerLysCysLeuAlaSerGlyTyrGly 158
seq_name: sp_rodent:Q9Z1R9
```

```

seq_documentation_block:
ID Q9Z1R9 PRELIMINARY; PRT; 246 AA.
AC Q9Z1R9;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE TRYPSINOGEN 16.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA ROWEN L., HOOD L.;
RT "Comparison between strains Balb/C and 129 in a region of the mouse T
RT cell receptor beta locus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF107342; AAC79093.1; -.
SQ SEQUENCE 246 AA; 26134 MW; 36625744 CRC32;
```

```

alignment_scores:
Quality: 243.00 Length: 115
Ratio: 2.793 Gaps: 2
Percent Similarity: 75.652 Percent Identity: 39.130
```

alignment_block:

US-09-030-606-177 x Q9Z1R9 ..

Align seg 1/1 to: Q9Z1R9 from: 1 to: 246

```

4 CACTCGACGCCCTGGCAGGCGGCACTGGTCATGAAAAAGCAATGTCTTG 53
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
33 AsnSerValProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCys 48
54 CTCGGGCGTCTGTGTCATCCGACAGTGGGTGTCTGACCCGACACTGTT 103
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
48 sGlyGlySerLeuIleAsnAspGlnTyrValValSerAlaAlaHisCyst 65
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 153
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
65 yRlyThrArgIleGlnValArgLeuGlyLysHisAsnIleAsnVal... 80
154 CAAGAGCCAGGAGCCAGATGTGTGAGGCCAGCCTTCCTACGCGACCC 203
||| :::|||||:::|||||:::|||||:::|||||:::|||||
81 LeuGluGlyAsnGluGlnPheIleAspAlaAlaLysIleIleLysHisPr 97
204 AGAGTACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGG 253
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
97 oAsnPheAsnArgLysThrLeuAsnAsnAspIleMetLeuIleLysLeus 114
254 ACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCATTGCTTGC 303
```



```
114 erSerProValThrLeuAsnAlaArgValAlaThrValAlaLeuProSer 130
      :::: ||::: ::::: ::::: ::::: ::::: ||
304 CAGTGCCCTACCGCGGGAAGCTCTGCTGCTTCTGCGCTGGGT 348
      ::||| ||||:::|||||:::|||||
131 SerCysAlaProAlaGlyThrGlnCysLeuIleSerGlyTrpGly 145
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:46 ; Search time 1809.22 Seconds
(without alignments)
1220.009 Million cell updates/sec

Title: us-09-030-606-177
Perfect score: 1119
Sequence: 1 GCGCACTCGCAGCCCTGCA.....ATGTTAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NUC
Searched: 2546578 segs, 986266752 residues

Database : EST:*

| | |
|-----|------------|
| 1: | em_est1:* |
| 2: | em_est2:* |
| 3: | em_est3:* |
| 4: | em_est4:* |
| 5: | em_est5:* |
| 6: | em_est6:* |
| 7: | em_est7:* |
| 8: | em_est8:* |
| 9: | em_est9:* |
| 10: | em_est10:* |
| 11: | em_est11:* |
| 12: | em_est12:* |
| 13: | em_est13:* |
| 14: | em_est14:* |
| 15: | em_est15:* |
| 16: | em_est16:* |
| 17: | em_est17:* |
| 18: | em_est18:* |
| 19: | em_est19:* |
| 20: | gb_est1:* |
| 21: | gb_est2:* |
| 22: | gb_est3:* |
| 23: | gb_est4:* |
| 24: | gb_est5:* |
| 25: | gb_est6:* |
| 26: | gb_est7:* |
| 27: | gb_est8:* |
| 28: | gb_est9:* |
| 29: | gb_est10:* |
| 30: | gb_est11:* |
| 31: | gb_est12:* |
| 32: | gb_est13:* |
| 33: | gb_est14:* |
| 34: | gb_est15:* |
| 35: | gb_est16:* |
| 36: | gb_est17:* |
| 37: | gb_est18:* |
| 38: | gb_est19:* |
| 39: | gb_est20:* |
| 40: | gb_est21:* |
| 41: | gb_est22:* |
| 42: | gb_est23:* |
| 43: | gb_est24:* |
| 44: | gb_est25:* |
| 45: | gb_est26:* |
| 46: | gb_est27:* |
| 47: | gb_est28:* |
| 48: | gb_est29:* |
| 49: | gb_est30:* |
| 50: | gb_est31:* |
| 51: | gb_est32:* |
| 52: | em_est20:* |
| 53: | em_est21:* |

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 361 | 32.3 | 777 | 48 | AI557281 | AI557281 PT2.1_15- |
| 2 | 223.6 | 20.0 | 722 | 48 | AI557025 | AI557025 PT2.1_10- |
| 3 | 214.2 | 19.1 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 4 | 205.8 | 18.4 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 5 | 170 | 15.2 | 216 | 32 | AA336074 | AA336074 EST40886 |
| 6 | 99.6 | 8.9 | 479 | 46 | AI436093 | AI436093 th92b12.x |
| 7 | 99.6 | 8.9 | 475 | 46 | AI452907 | AI452907 t146g04.x |
| 8 | 92 | 8.2 | 245 | 31 | AA295609 | AA295609 EST100780 |
| 9 | 91.6 | 8.2 | 264 | 32 | AA366980 | AA366980 EST77998 |
| 10 | 82.2 | 7.3 | 429 | 35 | AA594076 | AA594076 nn31a12.s |
| 11 | 81.4 | 7.2 | 407 | 38 | AA776018 | AA776018 ae83g11.s |
| 12 | 81 | 7.2 | 523 | 35 | AA542994 | AA542994 nj55b01.s |
| 13 | 76.2 | 6.8 | 486 | 35 | AA535837 | AA535837 nj79g05.s |
| 14 | 76.2 | 6.8 | 572 | 35 | AA579026 | AA579026 nf34f02.s |
| 15 | 74.8 | 6.7 | 253 | 20 | T29510 | T29510 EST82448 Hu |
| 16 | 74 | 6.6 | 546 | 33 | AA388907 | AA388907 mp16b03.r |
| 17 | 72.6 | 6.5 | 336 | 34 | AA506459 | AA506459 nh46a08.s |
| 18 | 72.6 | 6.5 | 548 | 35 | AA578976 | AA578976 nf26g03.s |
| 19 | 72.6 | 6.5 | 316 | 35 | AA593245 | AA593245 nn07e10.s |
| 20 | 72.6 | 6.5 | 902 | 47 | AI525832 | AI525832 PT1.3_06- |
| 21 | 72.6 | 6.5 | 732 | 48 | AI547309 | AI547309 PN001_AH- |
| 22 | 71.6 | 6.4 | 470 | 47 | AI525128 | AI525128 promrna-7 |
| 23 | 71 | 6.3 | 610 | 48 | AI557591 | AI557591 pt2.1-1.f |
| 24 | 70 | 6.3 | 371 | 34 | AA503943 | AA503943 nh38d05.s |
| 25 | 70 | 6.3 | 484 | 36 | AA603529 | AA603529 np15b08.s |
| 26 | 70 | 6.3 | 523 | 36 | AA639901 | AA639901 np08f03.s |
| 27 | 70 | 6.3 | 737 | 48 | AI547285 | AI547285 PN001_AH- |
| 28 | 70 | 6.3 | 454 | 50 | AI685510 | AI685510 tu36b10.x |
| 29 | 69.8 | 6.2 | 703 | 48 | AI546857 | AI546857 PN2.1_07- |
| 30 | 69.4 | 6.2 | 457 | 45 | AI385433 | AI385433 m185b05.y |
| 31 | 69.4 | 6.2 | 595 | 48 | AI557389 | AI557389 PT2.1_6-G |
| 32 | 69 | 6.2 | 619 | 26 | W58737 | W58737 zd23e07.r1 |
| 33 | 69 | 6.2 | 435 | 29 | AA177005 | AA177005 nc03b02.s |
| 34 | 68.8 | 6.1 | 432 | 38 | AA778667 | AA778667 af87e12.s |
| 35 | 67.8 | 6.1 | 340 | 27 | AA038537 | AA038537 m185b05.r |
| 36 | 67.2 | 6.0 | 519 | 42 | AI098337 | AI098337 vg84b03.r |
| 37 | 67.2 | 6.0 | 760 | 42 | AI127299 | AI127299 qb74f07.x |
| 38 | 66.8 | 6.0 | 453 | 34 | AA533652 | AA533652 nj72g08.s |
| 39 | 65.6 | 5.9 | 616 | 48 | AI546898 | AI546898 PN2.1_09- |
| 40 | 65.4 | 5.8 | 457 | 42 | AI078417 | AI078417 oz05e02.x |
| 41 | 65 | 5.8 | 624 | 47 | AI525001 | AI525001 promrna-4 |
| 42 | 64.2 | 5.7 | 397 | 35 | AA573604 | AA573604 nf42d12.s |
| 43 | 64 | 5.7 | 472 | 28 | AA062294 | AA062294 ml70a06.r |
| 44 | 64 | 5.7 | 482 | 28 | AA066464 | AA066464 ml73c10.r |
| 45 | 64 | 5.7 | 568 | 28 | AA068804 | AA068804 mm62g08.r |

ALIGNMENTS

RESULT 1

AI557281

LOCUS AI557281 777 bp mRNA

DEFINITION PT2.1_15_g12.r tumor2 Homo sapiens CDNA 3', mRNA sequence.

ACCESSION AI557281

NID g4489644

VERSION AI557281.1 GI:4489644

KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 777)
Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source Location/Qualifiers
1..777

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 32.3%; Score 361; DB 48; Length 777;
Best Local Similarity 100.0%; Pred. No. 6.9e-96;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCACTCGCAGCCCTGGCAGCGGGCACTGGTCATGGAACGAATTGTTCTGCTGGGC 60
Db 14 GCGCACTCGCAGCCCTGGCAGCGGGCACTGGTCATGGAACGAATTGTTCTGCTGGGC 73
QY 61 GTCCTGTGTCATCCGAGTGGTGTCTGTACGCCGACACTGTTCCAGAACTCCTACACC 120
Db 74 GTCCTGTGTCATCCGAGTGGTGTCTGTACGCCGACACTGTTCCAGAACTCCTACACC 133
QY 121 ATCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGGAG 180
Db 134 ATCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGGAG 193
QY 181 GCCAGCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATG 240
Db 194 GCCAGCTCTCCGTACGCGACCCAGAGTACACAGACCCTTGCTCGCTAACGACCTCATG 253
QY 241 CTCATCAAGTTGGAGCAATCCGTCGCCAGTCTGACACACCATCCGGAGCATCAGCATTTGCT 300
Db 254 CTCATCAAGTTGGAGCAATCCGTCGCCAGTCTGACACACCATCCGGAGCATCAGCATTTGCT 313
QY 301 TCGCAGTGGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTGGGGTCTGCTGCGGAAC 360
Db 314 TCGCAGTGGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTGGGGTCTGCTGCGGAAC 373
QY 361 G 361
Db 374 G 374

RESULT 2
AI557025 722 bp mRNA EST 23-MAR-1999
LOCUS AI557025 PT2.1.10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557025
NID 94489388
VERSION AI557025.1 GI:4489388

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 722)
Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source Location/Qualifiers
1..722

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 146 a 156 c 195 g 176 t 49 others
ORIGIN

Query Match 20.0%; Score 223.6; DB 48; Length 722;
Best Local Similarity 95.6%; Pred. No. 2.1e-55;
Matches 237; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 109 AACTCTACACCATCGGGCTGGGCTGACACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 168
Db 115 AGCTTCTACACCATCGGGCTGGGCTGACACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 174
QY 169 CAGATGTTGAGAGCCAGCTCTCCGTACGGCAGCCAGAGTACACAGACCTTGTGCT 228
Db 175 CAGATGTTGAGAGCCAGCTCTCCGTACGGCAGCCAGAGTACACAGACCTTGTGCTGCT 234
QY 229 AACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCGAGTCTGACACCATCCGAGC 288
Db 235 AACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCGAGTCTGACACCATCCGAGC 294
QY 289 ATCAGCATTTGCTTGCAGTGGCCCTACCGCGGGGAACTCTTGCC-TCGTTCTGGCTGGG 347
Db 295 ATCAGCATTTGCTTGCAGTGGCCCTACCGCGGGGAACTCTTNCCTTCTGCTGGG 354
QY 348 TCTGCTGG 355
Db 355 TCTGCTTG 362

RESULT 3
AA551449 415 bp mRNA EST 05-SEP-1997
LOCUS AA551449 n55e05.s1 NCI_GAP_P9 Homo sapiens cDNA clone IMAGE:996415
DEFINITION similar to SW:K1KA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 415) |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index. |
| JOURNAL | Unpublished (1997) |
| COMMENT | On May 9, 1995 this sequence version replaced g1:802243. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
Clone Sequencing by: Washington University Genome Sequencing Center
DNA distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html

Insert Length: 640 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.

FEATURES

1. .415

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_lib="NCI_CGAP_Prg"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
lab_host="DH10B"

```

/note="Organ: prostate; Vector: pAMP10; mRNA made from normal prostatic epithelial cells, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by

| | | | | |
|------------|------|-------|-------|------|
| BASE COUNT | 78 a | 125 c | 129 g | 83 t |
| ORIGIN | | | | |

| | | | | |
|-------------|--------|--------------|--------|-------------|
| Query match | 19.18; | Score 214.2; | DB 35; | Length 415; |
|-------------|--------|--------------|--------|-------------|

Best Local Similarity 98.6%; Pred. No. 1e-52;
Matches 216; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

| Qy | 143 | TTGAGGCGCCACCAAGAGCCAGGGAGGCCAGATGGTGGAGGCCAGCTCTCCGTACGGCACC | 202 |
|----|-----|---|-----|
| | | | |
| Db | 2 | TCGACGCGCCACCAAGAGCCAGGGAGGCCAGACGGTGGAGGCCAGCTCTCCGTACGGCACC | 61 |

[illegible]

0y 263 TGTCCAGTCTGACACCATCCGAGCATCAGATTGCTTGCAGTGCCTACCGCGGGG 322
 |||||
 122 TGTCCAGTCTGACACCATCCGAGCATCAGATTGCTTGCAGTGCCTACCGCGGGG 1811

QY 323 ACTCTTGCCCTGTTCTGGCTGGGGTCTGCTGGCGAACG 361
|||||
Db 182 ACTCTTGCCCTGTTCTGGCTGGGGTCTGCTGGCGAACG 220

| RESULT | 4 |
|------------|--|
| AI686689/c | |
| LOCUS | |
| DEFINITION | |
| AI686689 | 569 bp mRNA |
| tu55g11.x1 | NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2253092 3' |
| similar to | TR:Q92046 Q92046 PREPROTRYP SIN PRECURSOR ;, mRNA |
| sequence. | |

| | |
|-----------|--------------|
| ACCESSION | AI686689 |
| NID | 94897983 |
| VERSION | AI686689.1 |
| KEYWORDS | GI:4897983 |
| SOURCE | EST. |
| ORGANISM | human. |
| | Homo sapiens |

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 569)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CCGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced g1:3189584

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from G1bco
High quality sequence stop: 444.

FEATURES
source

1. .569

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2253092"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"

```

/note="Organ: prostate; Vector: PT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 125 a | 150 c | 170 g | 124 t |
| ORIGIN | | | | |

```

Query Match      18.48;  Score 205.8;  DB 50;  Length 569;

```

Best Local Similarity 94.7%; Pred. No. 3.4e-50;
Matches 213; Conservative 0; Mismatches 12; Indels 0; Gaps 0

Oy 157 GAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAACAGA 216
 |||||
 Db 569 GAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAACAGA 510

0y 217 CCCCTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAAATCCGTGTCCGAGTCTGAC 276
 |||||
 Db 509 CCCCTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAAATCCGTGTCCGAGTCTGAC 450C

QY 277 ACCATCCGAGCATCAGCATTCCTTCGAGTGCCCTACCGGGGAACTCTTGCTCGTT 336
 |||||
 Db 449 ACCATCCGAGCATCAGCATTCCTTCGAGTGCCCTACCGGGGAACTCTTGCTCGTT 390

QY 337 TCTGGCTGGGGCTGCTGGCGAAGATGCTGTGATTGCCATCCAG 381
|||||
Db 389 TCTGGCTGGGGCTGCTGGCGAAGCGCAGATGCCCTACCGTCTG 345

| | |
|------------|--|
| RESULT | 5 |
| AA336074 | |
| LOCUS | 216 bp mRNA |
| DEFINITION | EST40886 Endometrial tumor Homo sapiens CDNA 5' end similar to similar to kallikrein family, mRNA sequence. |
| ACCESSION | AA336074 |
| NID | 91988560 |

Db 278 AAGACGACCAATTACAGGGCTCTTCAAGATGCAGGTGCTGCCCTCACAAATCTGTTTCG 219
QY 790 CCTG-TTGTAGTGAAGGTGCGCCCTCTGAGCCTCCAGGGGTGTGACAGTC-ACA 847
Db 218 AAAGACATTGCTCAAGGGTATATCTTCGATCCTCCCAACTGTGTAGTAGCTTAAA 159
QY 848 ATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTT-AAATCCCTCATGCTCAGT 906
Db 158 GTGACTAATCCACTCCACCATCCCAATCTCCCTAAGTCTTGTGATTCCTCTCATATT 99
QY 907 ACACCAGGCGAGTCTAGCATTTTC 930
Db 98 AAACCAAGGAAGACCAAGCATTTTC 75

RESULT 7
AI452907/c 475 bp mRNA EST 13-APR-1999
LOCUS tJ46g04.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens CDNA clone
DEFINITION IMAGE:2144598 3' similar to TR:002710 002710 GAG POLYPEPTIDE. ;,
mRNA sequence.
ACCESSION AI452907
NID 94288636
VERSION AI452907.1 GI:4288636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 475)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949381.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 693 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 455.

FEATURES
source
1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19"
/clone="IMAGE:2144598"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 143 a 85 c 125 g 122 t
ORIGIN

Query Match 8.9%; Score 99.6; DB 46; Length 475;

Best Local Similarity 62.5%; Pred. No. 6,4e-19;
Matches 240; Conservative 0; Mismatches 134; Indels 10; Gaps 5;
QY 557 AGTCAGACTATCATGATTTACTGTGTGACTGTGCTGTCTATTTGACTAACCATGCC----- 613
Db 458 AGACAGACTATTCTGATTTGCTGTGCTTTGCCCTCTGCTGTCTATACAGAAAGCTACACCACC 399
QY 613 --GATGTTAGGTGAATTAAGCTCACTTGCCCTCAACCATCTTGTTATCCAGTTATCCT 670
Db 398 TTGCCCTTGTAGTGTGAGTGTCTGACCTTGCCACCTGACCTCAGATTCATATTTC 339
QY 671 CACTGAATGAGATTTCTCTGCTCACTGTGACGACCATCCAC-ATAATTTGACCTACA 729
Db 338 CATGTATTAAATTTTGTAGTGAAGTGAAGTGTGTTCCCACTGTTAGATCTGACATACC 279
QY 730 GAGTGAGGATCATATAGCTCTCAAGATGCTGTACTCCCTCACAAATTCATTTCT 789
Db 278 AAGACGACCAATTACAGGGCTCTTCAAGATGCAGGTGCTGCCCTCACAAATCTGTTTCG 219
QY 790 CCTG-TTGTAGTGAAGGTGCGCCCTCTGAGCCTCCAGGGTGGTGTGACAGTC-ACA 847
Db 218 AAAGACATTGCTCAAGGGTATATCTTCTGAATCTCTCCCACTTGTTGAGTAGCTTAAA 159
QY 848 ATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTT-AAATCCCTCATGCTCAGT 906
Db 158 GTGACTAATCCACTCCACCACCAATCTCCCTAAGTCTTTGATTCCTCTCATATT 99
QY 907 ACACCAGGCGAGTCTAGCATTTTC 930
Db 98 AAACCAAGGAAGACCAAGCATTTTC 75

RESULT 8
AA295609
LOCUS EST100780 Pancreas tumor I Homo sapiens CDNA 5' end similar to
DEFINITION kallikrein, pancreatic, mRNA sequence.
ACCESSION AA295609
NID 91947944
VERSION AA295609.1 GI:1947944
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 245)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulder,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
Small,R.V., Spriggs,T.A., Uterback,T.R., Weldman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280

TITLE
JOURNAL
MEDLINE
COMMENT
On Nov 29, 1993 this sequence version replaced gi:635027.
Other ESTs: THC167057
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA

[illegible]

| | | | |
|---|--|---|--------------------|
| Db | 16 | GCCTGTCTGTGGGGGTGCACACTGTGTCGCCGTGCCCTCATCTCAGTCTCGGATTGTGGAGGC | 75 |
| QY | 400 | TGGAGTGTGAGAGCTTTCCCAACCCCTGGCAGG----GTTGTACCATTTTCGGCAACTTC | 455 |
| Db | 76 | TGGAGTGTGAGAGCATTTCCCAACCCCTGGCAGGTGGCTGTGTACAGTCATGGATGGCA | 135 |
| QY | 456 | CAGTCAGAGACGCTCCTGCTGCATCTCTACTGGGTGCTCCTACTGCTCAGTCATCACC | 515 |
| Db | 136 | CACTGTGGGGGTGTCTGTGTGCACACCCAGTGGGTCTCAGCTGCCCCATTGCTTAAG | 195 |
| QY | 516 | CGGACA 522 | |
| Db | 196 | AAGAATA 202 | |
| RESULT | 15 | | |
| T29510 | | | |
| LOCUS | T29510 | 253 bp | EST 06-SEP-1995 |
| DEFINITION | EST82448 Human Prostate gland Homo sapiens cDNA 5' end similar to kallikrein 1, renal/pancreas/salivary (HT:711), mRNA sequence. | | |
| ACCESSION | T29510 | | |
| NID | 9611608 | | |
| VERSION | T29510.1 | GI:611608 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | |
| AUTHORS | Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinklejr,P.S., Kelley,J.M., Klimek,K.M., Kelley,J.C., Lau,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancues,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C. | | |
| TITLE | Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence | | |
| JOURNAL MEDLINE | Nature 377, 3-174 (1995) | | |
| COMMENT | 96026280 | | |
| CONTACT: | Venter, JC | | |
| INSTITUTE: | The Institute for Genomic Research | | |
| ADDRESS: | 932 Clopper Rd, Gaithersburg, MD 20878 | | |
| TEL: | 3018699056 | | |
| FAX: | 3018699423 | | |
| EMAIL: | tdbinfo@tdb.tigr.org | | |
| FOR CLONE AVAILABILITY, ADDITIONAL SEQUENCE AND EXPRESSION INFORMATION RELATED TO THIS EST, PLEASE CONTACT THE TIGR DATABASE (tdbinfo@tdb.tigr.org) | | | |
| SEQ PRIMER: | M13 Reverse. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..253 | | |
| ORGANISM="Homo sapiens" | | | |
| /db_xref="ATCC (lnhost):106507" | | | |
| /db_xref="taxon:9606" | | | |
| /clone_lib="Human Prostate gland" | | | |
| /note="Organ: prostate gland" | | | |
| BASE COUNT | 42 a 68 c 81 g 59 t 3 others | | |
| ORIGIN | | | |
| Query Match | 6.7% | Score 74.8; | DB 20; Length 253; |

Best Local Similarity 65.8%; Pred. No. 1e-11;
Matches 123; Conservative 0; Mismatches 60; Indels 4; Gaps 1;

```
QY 340 GGCTGGGGTCTGCTGGCGAAGATGCTGTGATGCCATCCAGTCCAGACTGTGGAGGC 399
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 GCCTGTCTGTGGGGTGCATGTGCGCGTGCCTCATCCAGTCTCGGATGTGGAGGC 98
QY 400 TGGGAGTGTGAGAAGCTTTCCCAACCCCTGGCAG---GTGTACCATTTGGCAACTTC 455
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99 TGGAGTGTGAGAAGCATTTCCCAACCCCTGGCAGTGTGTACAGTCATGGATGGCA 158
QY 456 CAGTCAAGGACGCTCTGTCATCCTCACTGGGTGCTCACTACTGTCACTGCATCACC 515
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 CACTGTGGGNTNTCTGTGACACCCCAAGTGGGTCTCACAGCTGCCCATGTGNTAAG 218
QY 516 CGGAACA 522
      ||| | |
Db 219 AAGAATA 225
```

Search completed: September 28, 1999, 11:29:50
Job time: 1871 sec

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th st SE,
Bothell, WA 98021, USA
location/Qualifiers
FEATURES
Source 1. .1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1. .765
/note="androgen regulated"
/codon_start=1
/product="serine protease prostate"
/protein_id="AAD21580.1"
/db_xref="PID:g4512030"
/db_xref="GI:4512030"
/translation="MATAGNPMWGMFLGYLLILGVAGSLVSGSCSQIINGEDCPHSQPM
QALVMELEDFCGGLVHPQWVLSAHCFOINSYITIGLHSLIADQEPGDSOMVEASLS
VRHPEYNRPILANDLMLIKLDESSESITRISISIASQCPITAGNSCLVSGWGLANGR
MPTVLOCVNVSVSEEVCSKLYDPLYPHPSMFCAGGHDQKDSCHGSDGGLICNGSLQ
GLVSFGKAPCGGVGPVYTNLCKFTWEIKTVQAS"
1327. .1332

polyA_site 269 a 489 c 334 g 255 t
BASE COUNT
ORIGIN

Query Match 32.2%; Score 360.8; DB 42; Length 1347;
Best Local Similarity 96.8%; Pred. No. 1.2e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 CGCACTCGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAAATGTTGCTCGGGCG 61
Db 116 CGCACTCGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAAATGTTGCTCGGGCG 175
QY 62 TCCTGTGCATCCGACAGTGGGTGCTGTACGCCGACACTGTTCCAGAACTCCTACACCA 121
Db 176 TCCTGTGCATCCGACAGTGGGTGCTGTACGCCGACACTGTTCCAGAACTCCTACACCA 235
QY 122 TCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGAGCCAGATGGTGAGG 181
Db 236 TCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGCCAGGAGCCAGATGGTGAGG 295
QY 182 CCAGCCTCTCCGTACGGCGCAGAGTACACACAGACCTTGCTCGCTAACGACCTCATGC 241
Db 296 CCAGCCTCTCCGTACGGCGCAGAGTACACACAGACCTTGCTCGCTAACGACCTCATGC 355
QY 242 TCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCGAGCATCATGCTT 301
Db 356 TCATCAAGTTGGACGAATCCGTGTCGAGTGTGACACCATCCGAGCATCATGCTT 415
QY 302 CGCAGTGCCTACCGCGGGGAACTCTGCTGCTTCTGGCTGGGCTGCTGGCGAAGC 361
Db 416 CGCAGTGCCTACCGCGGGGAACTCTGCTGCTTCTGGCTGGGCTGCTGGCGAAGC 475
QY 362 ATGCTGTGATGCCATCCAG 381
Db 476 GCAGATGCCTACCGTGTG 495

RESULT 2
SSU76256 1140 bp mRNA MAM 03-JAN-1998
LOCUS
DEFINITION Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,
complete cds.
ACCESSION U76256
NID 92737920
VERSION 076256.1 GI:2737920
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 1140)
AUTHORS Simmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,

TITLE Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
Purification, Characterization and Cloning of Enamel Matrix Serine
Proteinase 1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1140)
AUTHORS Simmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,
Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA

FEATURES
source 1. .1140
/organism="Sus scrofa"
/db_xref="taxon:9823"
69. .140
/note="the signal peptide is cleaved after Ala24 and the
preprotein is secreted into the developing enamel matrix"
69. .833
/note="EMSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAB94638.1"
/db_xref="PID:g2737921"
/db_xref="GI:2737921"
/translation="MTAGNPMWGMFLGYLLILGSLAFINGHITINGEDCPHSQPM
QALFLEDFECGGLVHPQWVLSAHCFOINSYITIGLHLEPEQEPGDSOMVEASLS
IOHPEYNRPSMANDLMLIKLDESYSLSITVRNISVVSQCPITGSDCLVSGWGLASGR
LPQVLOCVNVSVSEEVCSKLYDPLYPHPSMFCAGGHDQKDSCHGSDGGLICNGSLQ
GLVSFGQAQCGGPNVGPVYTNLCKFTDWIQTIGAS"
159. .830
/note="the preprotein is cleaved following His30
activating the serine proteinase; the active protein has
an apparent molecular weight of 34 kDa and a derived
molecular weight of 24 kDa"
/product="enamel matrix serine proteinase 1"
order(177. .179,549. .551)
/note="disulfide bond"
order(234. .236,282. .284)
/note="disulfide bond"
join(279. .281,414. .416,687. .689)
/note="encodes catalytic triad"
405. .407
/note="encodes potential glycosylation location; yields a
blank cycle during protein sequencing"
order(408. .410,789. .791)
/note="disulfide bond"
order(510. .512,705. .707)
/note="disulfide bond"
order(675. .677,750. .752)
/note="disulfide bond"

BASE COUNT 245 a 366 c 295 g 234 t
ORIGIN

Query Match 24.2%; Score 270.6; DB 3; Length 1140;
Best Local Similarity 84.9%; Pred. No. 3.6e-66;
Matches 303; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 2 CGCACTCGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAAATGTTGCTCGGGCG 61
Db 184 CCCACTCGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAAATGTTGCTCGGGCGG 243
QY 62 TCCTGTGCATCCGACAGTGGGTGCTGTACGCCGACACTGTTCCAGAACTCCTACACCA 121
Db 244 TCCTGTGCATCCGACATGGGTGCTGTACGCCGACACTGTTCCAAATTCCTACACCA 303
QY 122 TCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGTGAGG 181
Db 304 TCGGGCTGGGTCTGACAACTTTGAGGCCGGAACAAGACCAGCCAGATGATGAGG 363
QY 182 CCAGCCTCTCCGTACGGCGCAGAGTACACAGACCTTGCTGCTAACGACCTCATGC 241
|||||

Db 364 CCAGCCTCTCCATTACAGCACCAGAAATACACAGCCGCTCTATGGCCACGACCTCATGC 423

QY 242 TCATCAAGTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCATGCTT 301

Db 424 TCATCAAGTGAAGAATCGGTGTGCTGTGACACCGTCCGGAACATCAGCGTCTCT 483

QY 302 CGCAGTGCCTTACCGCGGGAAGTCTTGCTCGTTCTGGCTGGGCTCTGCTGGCGA 358

Db 484 CCCAGTGCCTTACCGCGGGAAGTCTTGCTCGTTCTGGCTGGGCTCTGCTGGCGA 540

RESULT 3

AF13141 5900 bp DNA PRI 25-MAR-1999

LOCUS Homo sapiens serine protease gene, complete cds.

DEFINITION AF13141

ACCESSION AF13141

NID 94512031

VERSION AF13141.1 GI:4512031

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.

AUTHORS Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.

TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)

MEDLINE 99179024

REFERENCE 2 (bases 1 to 5900)

AUTHORS Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA

FEATURES

source

1. 5900

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13"

join(<1411. .1471,2735. .2897,3319. .3569,3653. .3789, 5062. .5817)

/product="serine protease prostase"

1411. .1471

/number=1

join(1411. .1471,2735. .2897,3319. .3569,3653. .3789, 5062. .5214)

/note="androgen regulated"

/codon_start=1

/product="serine protease prostase"

/protein_id="AAD21581.1"

/db_xref="PID:94512032"

/db_xref="GI:4512032"

/translation="MATAGNPMWGLGILILGAGSLVSGSQIINGEDCSPHSQPM OALVMENELFCSGVLVHPQWVLSAHCFOINSYITIGLHSLLEADQEPGSQMVESLS VRHPEYNRPLANDLMLIKLDESSESPTIRISISIASQCPAGNSCLVSGWGLANGR MPTVLQCVNVSVSEVCSKLYDPLHPSMFCAGGHDQKDSNGDSGGLICNGYLO GLVSFGKAPCGQVGPVYTNLCKFTWIEKTVOAS"

2735. .2897

/number=2

3319. .3569

/number=3

3653. .3789

/number=4

5062. .5817

/number=5

1417 a 1643 c 1569 g 1271 t

BASE COUNT

ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 5900;

Best Local Similarity 98.8%; Pred. No. 1.1e-61;

Matches 257; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 105 CCAGACTCCTACACCATCCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGACCAGG 164

Db 3313 CCAGACTCCTACACCATCCGGGCTGGGCTGACAGTCTTGAGGCCGACCAAGACCAGG 3372

QY 165 GAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCT 224

Db 3373 GAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCT 3432

QY 225 CGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTACACCATCCG 284

Db 3433 CGCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTACACCATCCG 3492

QY 285 GAGCATCAGCATGCTGTGCGAGTGGCCCTACCGCGGGGAAGTCTTGCTCGTTCTGCGTG 344

Db 3493 GAGCATCAGCATGCTGTGCGAGTGGCCCTACCGCGGGGAAGTCTTGCTCGTTCTGCGTG 3552

QY 345 GGGTCTGCTGGCGAAGCATG 364

Db 3553 GGGTCTGCTGGCGAAGCATG 3572

RESULT 4

AF135023 4740 bp DNA PRI 20-APR-1999

LOCUS Homo sapiens kallikrein-like protein 1 KLK-L1 gene, partial cds.

DEFINITION AF135023

ACCESSION AF135023

NID 94589272

VERSION AF135023.1 GI:4589272

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Yousef, G.M., Luo, L.Y. and Diamandis, E.P.

AUTHORS Yousef, G.M., Luo, L.Y. and Diamandis, E.P.

TITLE Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4

JOURNAL Biochem. Biophys. Res. Commun. (1999) In press

REFERENCE 2 (bases 1 to 4740)

AUTHORS Yousef, G.M., Luo, L.Y. and Diamandis, E.P.

TITLE Direct Submission

JOURNAL Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

FEATURES

source

1. 4740

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="19"

/map="19q13.3-q13.4"

join(<2263. .2425,2847. .3097,3181. .3317,4588. .>4740)

/product="kallikrein-like protein 1 KLK-L1"

join(<2263. .2425,2847. .3097,3181. .3317,4588. .4740)

/codon_start=3

/product="kallikrein-like protein 1 KLK-L1"

/protein_id="AAD26424.1"

/db_xref="PID:94589273"

/db_xref="GI:4589273"

/translation="SLVSGSCSQIINGEDCSPHSQPMOALVMENELFCSGVLVHPQW VLSAHCFOINSYITIGLHSLLEADQEPGSQMVESLSVRHPEYNRPLANDLMLIKLD ESVSSEPTIRISISIASQCPAGNSCLVSGWGLANGRMPTVLQCVNVSVSEVCSKLYDPLHPSMFCAGGHDQKDSNGDSGGLICNGYLOGLVSFGKAPCGQVGPVYTN LCKFTWIEKTVOAS"

1106 a 1275 c 1308 g 1051 t

BASE COUNT

ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 4740;

Best Local Similarity 98.8%; Pred. No. 1e-61;


```

mat_peptide      145 . 816 /note="encodes propeptide"
BASE COUNT      289 a 386 c 286 g 276 t
ORIGIN

```

| | | | | |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match | 19.08; | Score 212.8; | DB 12; | Length 1237; |
| Best Local Similarity | 74.48; | Pred. No. 7.9e-50; | | |
| Matches 268; Conservative | 0; | Mismatches 92; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| OY | 2 | CGCACCTCGCAGCCCTGGCAGGCGGCCACTGTGCATGGAAACCAATTTCTGCTCGGGCG | 61 |
| Dd | 170 | CACACTCGCAACCCTTGGCAGGCGGCCACTGTTCCTCAGAAGACGGTTTTTTCTGCTCGGGAG | 229 |
| OY | 62 | TCCGTGTGCATCCGCCAGTGGGTGTCTGTACGCCGCACACTGTTTCCGAACCTCCTACACA | 121 |
| Dd | 230 | TCTTGTGTCATCCGCAGTAGTGGGTGTCTGTACGACGACACATTGCTTAAGAGAGTCTACATCG | 289 |
| OY | 122 | TCGGGCTGGGECTGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGTGGAG | 181 |
| Dd | 290 | TGGAGCTGGGCTGCATAAACCCTGAAAGGCTCCCAAGAGCCTGGCACCGGATGTAGAG | 349 |
| OY | 182 | CCAGCCTCTCCGTACGGCACCCAGAGTACAAAGACCCTTGCTGCTAACGACCTCATGC | 241 |
| Dd | 350 | CCCACCTCTCCATCCAGCACCCCCAACCTCAACGATCCTCTTTGCCAAACGATCCTCATGC | 409 |
| OY | 242 | TCATCAAGTTGACGCAATCCGTGTCCGAGTCTGACACCATCCGSGACATCAGATTGCTT | 301 |
| Dd | 410 | TCATCAAACCTGACGAGAGTCAAGTCAAGTCTTAACACTATCAGGAGCATCCCTGTGGCTA | 469 |
| OY | 302 | GCGAGTGCCTTACCCGCGGGGAACCTTTGCTCTGTGCTGGGGTCTGCTGGCGAAG | 361 |
| Dd | 470 | CCCAATGCCCACTCTTGAGATACCTGCTAGTCTCTGTGGTTGGGGTCAACTAAAGATG | 529 |

RESULT 7

| LOCUS | 199016 bp | PRI | 08-OCT-1998 |
|------------|--|-----|-------------|
| DEFINITION | Human DNA sequence from clone 48G12 on chromosome Xq27.1-27.3. | | |
| | Contains STRs and GSSs, complete sequence. | | |

| | |
|-----------|-----------------------|
| ACCESSION | AL031054 |
| NID | 93550059 |
| VERSION | AL031054.1 GI:3550059 |

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 199016)

| | |
|---------|--|
| AUTHORS | Pearce, A. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (21-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire |

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 48G12. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrx>

48G12 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

location/qualifiers
1. .199016

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="48G12"
/map="q27.1-27.3"
/clone_11b="RPC11"
11..1153
/note="MUT2_internal repeat: matches 4550. .5668 of
consensus"
repeat_region
1155..1450
/note="MUT2CA repeat: matches 1. .300 of consensus"
repeat_region
1425..1544
/note="MUT2CB repeat: matches 316. .444 of consensus"
repeat_region
1736..1779
/note="22 copies 2 mer at 93% conserved"
repeat_region
1819..1884
/note="MUT1D repeat: matches 288. .353 of consensus"
repeat_region
1914..1943
/note="15 copies 2 mer at 100% conserved"
repeat_region
2061..2100
/note="20 copies 2 mer at 83% conserved"
repeat_region
2138..2173
/note="MUT2FA repeat: matches 416. .450 of consensus"
repeat_region
complement(2636..2933)
/note="Alu5g repeat: matches 299. .2 of consensus"
repeat_region
complement(3141..3601)
/note="MUT1D repeat: matches 482. .1 of consensus"
misc_feature
3857..4383
/note="match: GSS B59439"
repeat_region
3901..3962
/note="FLAM repeat: matches 71. .132 of consensus"
repeat_region
3967..4000
/note="17 copies 2 mer aa 82% conserved"
repeat_region
complement(4206..4385)
/note="L1PA4 repeat: matches 893. .714 of consensus"
repeat_region
5064..5354
/note="L1MA7 repeat: matches 767. .1036 of consensus"
repeat_region
complement(7448..8028)
/note="L1MC1 repeat: matches 1074. .482 of consensus"
repeat_region
complement(8178..8370)
/note="L1MC1 repeat: matches 491. .298 of consensus"
repeat_region
complement(8371..9258)
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region
complement(9112..11605)
/note="L1 repeat: matches 5390. .2936 of consensus"
repeat_region
complement(11602..11892)
/note="L1MC1 repeat: matches 301. .5 of consensus"
repeat_region
complement(11750..12175)
/note="L1 repeat: matches 5390. .4955 of consensus"
repeat_region
12180..12479
/note="AluY repeat: matches 1. .299 of consensus"
repeat_region
complement(12527..13544)
/note="L1 repeat: matches 4944. .3919 of consensus"
repeat_region
complement(13756..13969)
/note="MIR repeat: matches 249. .30 of consensus"
repeat_region
complement(14025..14226)
/note="MIR repeat: matches 262. .52 of consensus"
repeat_region
15048..15735
/note="L1 repeat: matches 3616. .4317 of consensus"
repeat_region
16545..17275
/note="L1ME2 repeat: matches 150. .911 of consensus"
repeat_region
complement(19351..19472)
/note="MIR repeat: matches 253. .124 of consensus"
repeat_region
complement(19718..19889)
/note="L1 repeat: matches 5189. .5008 of consensus"
repeat_region
19890..20010
/note="AluJo repeat: matches 1. .124 of consensus"
complement(20041..20304)

```



```

repeat__region /note="L1 repeat: matches 5015. .4754 of consensus"
                complement(20316. .20556)
                /note="L1MA4A repeat: matches 1047. .801 of consensus"
                complement(20602. .20758)
repeat__region /note="L1 repeat: matches 2529. .2371 of consensus"
                complement(20749. .21857)
                /note="L1 repeat: matches 2203. .1103 of consensus"
                complement(21324. .22281)
                /note="MER25 repeat: matches 2136. .1177 of consensus"
                complement(22277. .22722)
repeat__region /note="MER25 repeat: matches 1031. .582 of consensus"
                complement(22807. .22998)
                /note="MER25 repeat: matches 205. .13 of consensus"
                complement(23002. .23044)
repeat__region /note="MSTA repeat: matches 384. .426 of consensus"
                complement(23045. .23443)
                /note="MST-INTERNAL repeat: matches 1. .394 of consensus"
                complement(23519. .23679)
repeat__region /note="L1MB1 repeat: matches 914. .757 of consensus"
                complement(23755. .23876)
                /note="match: STS L41276"
                complement(23842. .24139)
repeat__region /note="AluSc repeat: matches 1. .299 of consensus"
                complement(24613. .24899)
                /note="AluDb repeat: matches 293. .7 of consensus"
                complement(25248. .25604)
repeat__region /note="THE1B repeat: matches 364. .1 of consensus"
                complement(25605. .26715)
                /note="THE1B-INTERNAL repeat: matches 1580. .471 of
                consensus"
                complement(26728. .26769)
repeat__region /note="THE1C repeat: matches 44. .3 of consensus"
                complement(27295. .27591)
                /note="AluSq repeat: matches 299. .1 of consensus"
                complement(28402. .28571)
                /note="MER5B repeat: matches 174. .8 of consensus"
                complement(28721. .28894)
repeat__region /note="MIR repeat: matches 82. .258 of consensus"
                complement(29430. .30796)
                /note="L1 repeat: matches 4045. .5390 of consensus"
                complement(30659. .31555)
                /note="L1PA16 repeat: matches 1. .904 of consensus"
                complement(31625. .31648)
                /note="12 copies 2 mer tt 100% conserved"
                complement(31650. .31950)
repeat__region /note="AluYa5 repeat: matches 301. .1 of consensus"
                complement(31962. .32114)
                /note="L1MD1 repeat: matches 971. .806 of consensus"
                complement(31972. .32025)
                /note="MER42c repeat: matches 1538. .1485 of consensus"
                complement(32061. .33135)
repeat__region /note="MER42c repeat: matches 1301. .173 of consensus"
                complement(33214. .33525)
                /note="L1MC3 repeat: matches 1362. .1054 of consensus"
                complement(35486. .36087)
                /note="L1MA8 repeat: matches 617. .3 of consensus"
                complement(35931. .36173)
                /note="L1 repeat: matches 5390. .5158 of consensus"
                complement(36176. .36573)
                /note="L1MA9 repeat: matches 620. .1016 of consensus"
                complement(38289. .38580)
repeat__region /note="AluY repeat: matches 292. .1 of consensus"
                complement(38584. .40069)
                /note="L1 repeat: matches 4630. .3154 of consensus"
                complement(40030. .40877)
                /note="L1 repeat: matches 3086. .2235 of consensus"
                complement(40843. .40929)
                /note="L1MA3 repeat: matches 148. .61 of consensus"
                complement(40909. .41341)
repeat__region /note="L1 repeat: matches 5287. .4843 of consensus"
                complement(44097. .44291)
                /note="MIR repeat: matches 1. .204 of consensus"
                complement(44511. .44801)

```

```

repeat_region      /note="AluSx repeat: matches 285. .1 of consensus"
                    45620. .45721
                    /note="MIR repeat: matches 81. .189 of consensus"
repeat_region      complement(45926. .46063)
                    /note="MLTIF repeat: matches 541. .379 of consensus"
repeat_region      complement(47019. .47161)
                    /note="MIR repeat: matches 262. .120 of consensus"
repeat_region      47339. .47412
                    /note="MIR repeat: matches 64. .137 of consensus"
repeat_region      47884. .48136
                    /note="L1PA4 repeat: matches 643. .891 of consensus"
repeat_region      complement(48404. .48890)
                    /note="L1PA3 repeat: matches 893. .406 of consensus"
repeat_region      49119. .49216
                    /note="MIR repeat: matches 146. .246 of consensus"
misc_feature       complement(50414. .50631)
                    /note="match: STS G03807"
repeat_region      complement(51021. .51433)
                    /note="MLTID repeat: matches 505. .101 of consensus"

```

| | | | | |
|---------------------------|--------|--------------------|-----------|----------------|
| Query Match | 12.4%; | Score 138.4; | DB 9; | Length 199016; |
| Best Local Similarity | 62.2%; | Pred. No. 2.1e-28; | | |
| Matches 270; Conservative | 0; | Mismatches 156; | Indels 8; | Gaps 3; |

QY 505 ACTGCATCACC CGGAACACTGTGATCAACTAGCCAGCACCATAGTTCTCCGAAGTCAGAC 564

Db 136146 ATTTCACTTCTAGGACACTGTAATTAATTACCAAGCCAGAGCTCTACCTGAGTCAGAC 136087

QY 565 TATCATGATTACTGTGTTGACTGTGCTCTATTGTACTAACCATGCCGATGTTAGGTG 624

Db 136086 TATTCTGATTGCTGCTTTGGCCTCTGCTGCTATATTATGGTAGCTACGCCACCCTTGGTTT 136027

QY 625 AA-----ATTAGCTCACTGGCCTCAACCATCTGGTATCCAGTTATCCTCACTGAAT 678

Db 136026 GACAGTTGAGTGCTGCACTTGGCCACTGGCAGTTGGGATCCAATCATTCCTTGTAT 135967

679 TGAGATTTCCTGCTCAGTGTACGCCATTCCACATAATTCTGACCTACAGAGGTGAGG 738

Db 135966 TTAATTTTGAAGTTCAGTGCAGTTCACATGTTAGATCTGACATACAGAGAAGAGC 135907

QY 739 GATCATATAGCTCTTCAAGGATGCTGTTACTCCCTCACAAATTCAATTTCTCTCTGTTGA 798

Db 135906 AATTACAGGGCTCTTCAAGAATGCAGGTGCTGCCCTCAAAATCTATTTCAGAAGTATT 1358477

QY 799 GTGAAGGTGCGCCCTCTGGAGCCTCCCAAGGTGGGTGTGCAAGTC-ACATGATGATG 857

Db 135846 GGTCAAGGTTATCTTCTGGACCTCTCAGCTGGGATGAATAGGCTTAAGTGACTAATG 135787

QY 858 TATGATCGTGTTCACCATTAACCAAGCCTTAATCCC-TCATGCTCAGTACACACGAGGC 916

Db 135786 TATTCACCATCCCAATCTCCCTAAGCCTTTGGATCCCTTCCCTTTACATAAAACCAAGG 135727

QY 917 AGGCTAGCATTTC 930

Db 135726 AGATCAGGCATTTC 135713

RESULT 8
HSU76377

| | | | | | |
|------------|---|----------|-----|-----|-------------|
| LOCUS | HSU/b3// | 421/9 bp | BNA | PRI | 02-JAN-1999 |
| DEFINITION | Human olfactory receptor 17-93 (OR17-93) and olfactory receptor | | | | |

17-201 (OR17-201) genes, complete cds.
 F76377

NID 94098234
VERSION 176377 1
GT: A098234

KEYWORDS
SOURCE
human
.

ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Vertebrata: Mammalia:

Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 43170)

AUTHORS
Ferraz, C. and Demaille, J.G.

[illegible]

| QY | 847 | 11358 | QY | 906 | Db | 11418 | RESULT | 9 |
|------------|--|--|-------|-----|-------------|-------|--------|---|
| Db | 11298 | ACAAAGTATTAAATCAAGGGTATATCTTCTGGGCCCTCCACAGCTGGATGAGTAGTCTAA | 11357 | | | | | |
| QY | 847 | AATGATGAATGTATGATCGTGTCCCATTTACCCAAAGCCTTTAAAT-CCCTCATGCTCAG | 905 | | | | | |
| Db | 11358 | AGTGACTAATCCACATCCACATCCCTATCTCCCTAGCCCTTGATCTCCTCTACAT | 11417 | | | | | |
| QY | 906 | TACACCAGGCGCAGGCTTAGCATTTTC | 930 | | | | | |
| Db | 11418 | TAAACCAAGAAGATCAGGCATTTC | 11442 | | | | | |
| LOCUS | HSU78308 | 39038 bp | DNA | PRI | 02-JAN-1999 | | | |
| DEFINITION | Human olfactory receptor olfr17-201-1 (OR17-201-1) gene, olfactory receptor olfr17-32 (OR17-32) gene and olfactory receptor pseudo_olfr17-01 (OR17-01) pseudogene, complete cds. | | | | | | | |
| ACCESSION | U78308 | | | | | | | |
| NID | g4098463 | | | | | | | |
| VERSION | U78308.1 | GI:4098463 | | | | | | |
| KEYWORDS | human. | | | | | | | |
| SOURCE | human. | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | | | | | |
| AUTHORS | 1 (bases 1 to 39038) | | | | | | | |
| TITLE | Ferraz, C. and Demaille, J.G. | | | | | | | |
| JOURNAL | Sequence of cosmid ICRF105CE06173 of human chromosome 17p13.3 | | | | | | | |
| REFERENCE | olfactory receptor gene cluster, containing genes OR17-32 and OR17-201-1, and pseudogene OR17-01 | | | | | | | |
| AUTHORS | Unpublished | | | | | | | |
| TITLE | 2 (bases 1 to 39038) | | | | | | | |
| JOURNAL | Ferraz, C. and Demaille, J.G. | | | | | | | |
| REFERENCE | Direct Submission | | | | | | | |
| AUTHORS | Submitted (15-NOV-1996) CRBM, CNRS, 1919, route de Mende, PO Box 5051, Montpellier 34033, France | | | | | | | |
| TITLE | Location/Qualifiers | | | | | | | |
| JOURNAL | 1. 39038 | | | | | | | |
| REFERENCE | /organism="Homo sapiens" | | | | | | | |
| AUTHORS | /db_xref="taxon:9606" | | | | | | | |
| TITLE | /chromosome="17" | | | | | | | |
| JOURNAL | /map="17p13.3" | | | | | | | |
| REFERENCE | /clone="ICRF105CE06173" | | | | | | | |
| AUTHORS | complement(9530..10495) | | | | | | | |
| TITLE | /gene="OR17-201-1" | | | | | | | |
| JOURNAL | complement(9530..10495) | | | | | | | |
| REFERENCE | /gene="OR17-201-1" | | | | | | | |
| AUTHORS | /codon_start=-1 | | | | | | | |
| TITLE | /product="olfactory receptor olfr17-201-1" | | | | | | | |
| JOURNAL | /protein_id="AAD00277.1" | | | | | | | |
| REFERENCE | /db_xref="PID:g4098465" | | | | | | | |
| AUTHORS | /db_xref="GI:4098465" | | | | | | | |
| TITLE | /translation="MSLQKLMPEACNTNRATAVAEFIGLVQTEEMQPVFVLLFAYLVITIGNLSILAALVEPKLHAPYFELGNLSVLDVGCITVVPAMGRLLSHKSTISYDACLSQLEFFHLLAGMDCFLTLTAAYDRLLAICPLTYSTRSQTVQRLVAASWACAFETNALHTHTVAMSTLNFCEGPNVNHFCYCDLPOLQSCSSTQNELLLFVAAFMAVAPLVFISVSYAHVAAVLAQIRSAEGRKAFSTGSHLTVGIFGTGVFSYMRIGSVESSDKDKGVGVEMTVINPMLNPLIYSLRNTDVQALCQLLVGERSLT" | | | | | | | |
| FEATURES | 26023..26967 | | | | | | | |
| SOURCE | /gene="OR17-32" | | | | | | | |
| LOCUS | 26023..26967 | | | | | | | |
| DEFINITION | /gene="OR17-32" | | | | | | | |
| ACCESSION | /codon_start=-1 | | | | | | | |
| NID | /product="olfactory receptor olfr17-32" | | | | | | | |
| VERSION | /protein_id="AAD00276.1" | | | | | | | |
| KEYWORDS | /db_xref="PID:g4098464" | | | | | | | |
| SOURCE | /db_xref="GI:4098464" | | | | | | | |
| ORGANISM | /translation="MMGONQTSISDFLLGLPIOPEQONLCVALFLAMYLTTLLGNLLIYLRDLSHLHTPMYLFSLNLSFSDLCFSSVTIPKLLQNMQNDPSIPYADCLTOMYFLEGLDLESEFLVAMAYDRYAICFPLHYTAIMSPMCLTYALSWSLTFHAMLHTLMARLCFCADNVIPHEFCDSMLKLAFSDTRVNEWIFIMGILLVIFPLILGSYARIYSSILKVPSSKIGICKAFSTGSHLSVSLFYGTGIVIGLYLCCSSANSSTLKDTVMAMYTATVPMNLNPLIYSLRNRDMKALSRIVHOKTFFSL" | | | | | | | |

| | | | |
|---------------------------------|--|--|-----------------|
| gene | | complement(37565..38603) | /gene="OR17-01" |
| CDS | | /pseudo complement(37565..38603) /gene="OR17-01" /note="olfactory receptor pseudo_olfr17-01" /codon_start=1 /pseudo | |
| BASE COUNT | 11773 a 7672 c 7673 g 11920 t | | |
| ORIGIN | | | |
| Query Match | 12.0%; Score 134.2; DB 11; Length 39038; | | |
| Best Local Similarity | 65.4%; Pred. No. 2.4e-27; | | |
| Matches | 291; Conservative 0; Mismatches 143; Indels 11; Gaps 6 | | |
| QY | 497 TACTGCTCACTGCATCACC CGGAACACTGTGCATCACTAGCCAGCACCATAGTTCT-CCG 555 | | |
| Db | 4783 TCCTGTCAGTTTCACTCTTAGAACAACCATGATTAAITTAGCCAATGCCAGAGCTCTACAC 4724 | | |
| QY | 556 AAGTCAGACTATCATGATTACTGTGTGAGCTGTGCTGTCTATTGTACTAACCATGGCGAT 615 | | |
| Db | 4723 AAGTCAGACTATTCTGATGTGCTGTGTTCCTCTGCTGTGTCATTATGTGGTGACAGCCAA 4664 | | |
| QY | 616 GTTTAGGTGAA-----ATTAGCGTCACTTGGCCCTCAACCATCTTGTAATCCAGTTATCC 669 | | |
| Db | 4663 CTTCGCTTTTAATGTTGAGTGTGCTGCTGACCTTGGCCCCCTGCGACCTCGGATCCATTAATTC 4604 | | |
| QY | 670 TCACGTGAATTGAGATTTCCTGCTTCACTGTCAGGTCAGCCATTCCAC-ATAATTTCTGACCTAC 728 | | |
| Db | 4603 CCACTGCATTTAAATTTTATAGTTGAGTGAAGTGCAGTTCACCTGTAGATCTGCACATAC 4544 | | |
| QY | 729 AGAGTGAGGAGATCATATAGCTCTTCAAGATGCTGTACTCCCCTCACAAATTCATTTC 788 | | |
| Db | 4543 AGAGAGGAGACATCATAGAGCTCTTGAGAGTGCAGGTGCTGCCCTCACAAATTCATTTC 4484 | | |
| QY | 789 TC-CTGTTGTAGTGAAGGTGCGCCCTCTGAGACCTCCACAGGGTGCGTGCAGGTC-AC 846 | | |
| Db | 4483 ACAAGATTAATTAATCAAGGGTATATCTCTGCGCCCTCCACAGCTGGGATGAGTAGGTCTAA 4424 | | |
| QY | 847 AATGATGAATGTATGATCGTGTTCCTCAATACCAAGCCCTTAAT-CCCTCATGCTCAG 905 | | |
| Db | 4423 AGTGAATAATCCACCTCCAGCATCCCTATCTCCTTAAGCCCTTGATTCCCTCTACAT 4364 | | |
| QY | 906 TACACCAGGCGCAGGTCTAGCATTTTC 930 | | |
| Db | 4363 TAAACCAAAGAAGATCAGGCATTTTC 4339 | | |
| RESULT 10 | | | |
| AC007194_0/c | | | |
| WPCOMMENT | | | |
| Sequence split into 5 fragments | LOCUS AC007194 Accession AC007194 | | |
| Fragment Name | Begin | End | |
| AC007194_0 | 1 | 110000 | |
| AC007194_1 | 100001 | 210000 | |
| AC007194_2 | 200001 | 310000 | |
| AC007194_3 | 300001 | 410000 | |
| AC007194_4 | 400001 | 428051 | |
| LOCUS | AC007194 428051 bp DNA HTG 02-APR-1999 | | |
| DEFINITION | Homo sapiens chromosome 17 map 17p13.3, WORKING DRAFT SEQUENCE, in | | |
| | ordered pieces. | | |
| ACCESSION | AC007194 | | |
| NID | 94567182 | | |
| VERSION | AC007194.1 GI:4567182 | | |
| KEYWORDS | HTG; HTGS_PHASE2. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | |
| AUTHORS | Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 428051) Glusman,G., Sosinsky,A., Ben-Asher,E., Avidan,N., Sonkin,D., Bahar,A., Rosenthal,A., Clifton,S., Roe,B., Ferraz,C., Demallie,J. | | |

| | | | | | | |
|-----------------------|---|------------------------|--|-----------------|-------------|---------|
| TITLE | and Lancet,D. | | | | | |
| JOURNAL | Sequence, structure and evolution of a 430 kb human olfactory receptor gene cluster | | | | | |
| REFERENCE | Unpublished | | | | | |
| AUTHORS | 2 (bases 1 to 428051) | | | | | |
| | Glusman,G., Sosinsky,A., Ben-Asher,E., Avidan,N., Sonkin,D., Bahar,A., Rosenthal,A., Clifton,S., Roe,B., Ferraz,C., Demallle,J. and Lancet,D. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (02-APR-1999) Molecular Genetics and the Crown Human Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel | | | | | |
| COMMENT | * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| source | 1..428051 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17p13.3" | | | | | |
| BASE COUNT | 129870 a | 86849 c | 86715 g | 122459 t | 2158 others | |
| ORIGIN | | | | | | |
| Query Match | 12.0%; | Score 134.2; | DB 35; | Length 110000; | | |
| Best Local Similarity | 65.4%; | Pred. No. 2.8e-27; | | | | |
| Matches | 291; | Conservative | 0; | Mismatches 143; | Indels 11; | Gaps 6; |
| OY | 497 | TACTGCTCACTGCATCACC | CGGAACACTGTGATCACTAGCCAGCACCATAGTCT-CGG | 555 | | |
| Db | 27844 | TCCTGTCAGTTTCACCTCT | CINAGAACAACCATGATTATAGCCAATGCCAGAGCTCTACAC | 27785 | | |
| OY | 556 | AAGTCAGACTATCATGATACT | GTGTGACTGTGCTGTCTATTGTACTAACCATGCCGAT | 615 | | |
| Db | 27784 | AAGTCAGACTATTCTGATTGCT | GTGTTTGCCCTGCTGTCATTATGTTGGCTACAGCCA | 27725 | | |
| OY | 616 | GTTTAGGTGA----- | -ATTAGCGTCACTTGCCCTCAACCATCTTGTAACCATATCC | 669 | | |
| Db | 27724 | CTGTCTTTAATGTTGAGTGCT | GCACACTTGCCCCGCGACCTCGGATCCAAATTAATC | 27665 | | |
| OY | 670 | TCACCTGAATTGAGATTTCCT | CGCTTCAGTGTACGCCAATCCAC-ATAATTCTGACCTAC | 728 | | |
| Db | 27664 | CCACTGCATTTAAATTTTAAT | GTAGTGAAGTCACTGCAGTCCCACTGTAAGATCTGACATAC | 27605 | | |
| OY | 729 | AGAGGTGAGGATCATATAGCT | CTTCAAGGATGCTGTACTCCCCTCACAAATTCATTTC | 788 | | |
| Db | 27604 | AGAGAGGAGCATCATAGAGCT | CTTTGAGGATGCAGGTGCTGCCCTCACAAATCTATTTC | 27545 | | |
| OY | 789 | TC-CTGTTGTAGTGAAGGTG | CGCCCTCTGGAGCCCTCCCAAGGGTGGGTGTCAGGTC-AC | 846 | | |
| Db | 27544 | ACAAAGTATTATCAAGGGTAT | ATCTTCTTGCGCCCTCCAGCTGGGATGAGTAGTGTAA | 27485 | | |
| OY | 847 | AATGATGAATGTATGATCGTG | TCCCATTAACCCCAAAGCCTTTAAAT-CCCTCATGCTCAG | 905 | | |
| Db | 27484 | AGTACTAATTCACACTCCACG | ATCCCTATCTCCCTAAGCCTTTGATTCCTTCCTCTACAT | 27425 | | |
| OY | 906 | TACACCAGGGCAGGCTACGAT | TTTC | 930 | | |
| Db | 27424 | TAAACCAAGAAGATCAGCAT | TTTC | 27400 | | |
| RESULT | 11 | | | | | |
| LOCUS | HS15D7/c | 61450 bp | DNA | PRI | 11-DEC-1998 | |
| DEFINITION | Human DNA sequence from clone 15D7 on chromosome 6p22.1-22.3. Contains part of a 60S Ribosomal Protein L2 (L8) pseudogene, ESTs, STSs and GSSs, complete sequence. | | | | | |
| ACCESSION | NID | AL031229 | | | | |
| VERSION | AL031229.1 | GI:3947784 | | | | |
| KEYWORDS | HTG; 60S Ribosomal Protein L2; 60S Ribosomal Protein L8. | | | | | |

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 61450)
AUTHORS Phillips, S.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Dec 2, 1998 this sequence version replaced gi:3676174.
requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 15D7. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 34I8 (AL021918) is at 61351 in this
sequence. The true right end of clone 97D16 (AL009179) is at 100 in
this sequence. This sequence has been finished according to
sequence map criteria as follows. An attempt is made to resolve all
necessarily within known annotated human repeat sequence elements
(e.g. Alu). Where the sequence is ambiguous, there is an annotation
using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration by the Sanger
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,
David Ruddy, Jeffrey Gruen. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6/>
15D7 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2>.
FEATURES
Location/Qualifiers
1..61450
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="15D7"
/map="p22.1-22.3"
/clone_lib="RPC11"
381..821
/note="match: GSS AQ213610"
643..950
/note="AluSp repeat: matches 1..309 of consensus"
1312..1385
/note="tRNA-Arg-CGC repeat: matches 1..74 of consensus"
1369..1486
/note="match: STS Z78001"
complement(1494..1797)
/note="match: STS G02930"
1960..2001
/note="21 copies 2 mer ag 76% conserved"
1968..2003
/note="9 copies 4 mer agac 92% conserved"
2803..3107
/note="AluSq repeat: matches 1..308 of consensus"
3292..3366
/note="tRNA-Ile-ATT repeat: matches 1..75 of consensus"
3673..3730
/note="MER66A repeat: matches 284..340 of consensus"
5345..5643
/note="AluX repeat: matches 10..310 of consensus"
5751..5899
/note="FLAM_A repeat: matches 1..125 of consensus"
5903..5954
/note="26 copies 2 mer tg 79% conserved"
5927..5954
/note="7 copies 4 mer tgtg 100% conserved"

misc_feature
6100..6268
/note="match: GSS B55592"
6208..6511
/note="AluSq repeat: matches 1..296 of consensus"
6564..6804
/note="AluY repeat: matches 1..301 of consensus"
7108..7183
/note="tRNA-Phe-TTC repeat: matches 1..73 of consensus"
7205..7250
/note="23 copies 2 mer tt 96% conserved"
7466..7791
/note="AluSp repeat: matches 1..313 of consensus"
8151..8461
/note="AluY repeat: matches 3..304 of consensus"
8462..8579
/note="MER92C repeat: matches 200..311 of consensus"
8580..8664
/note="MER68A repeat: matches 1..84 of consensus"
8785..8920
/note="MER92C repeat: matches 414..552 of consensus"
9471..9769
/note="AluX repeat: matches 1..300 of consensus"
9773..9804
/note="16 copies 2 mer tt 91% conserved"
9882..10170
/note="AluJo repeat: matches 1..297 of consensus"
10186..10368
/note="L1MC2 repeat: matches 6158..6325 of consensus"
10370..10661
/note="AluSg repeat: matches 1..307 of consensus"
10676..10978
/note="AluSg repeat: matches 1..308 of consensus"
10991..11687
/note="L1MC2 repeat: matches 5173..5870 of consensus"
11682..12335
/note="L1MCA repeat: matches 663..1295 of consensus"
12236..12548
/note="AluSx repeat: matches 1..311 of consensus"
12549..12862
/note="L1MCA repeat: matches 371..663 of consensus"
12937..13246
/note="AluY repeat: matches 1..309 of consensus"
13294..13389
/note="L1MD2 repeat: matches 6242..6330 of consensus"
13390..13687
/note="AluSx repeat: matches 12..310 of consensus"
13688..13916
/note="L1MD2 repeat: matches 6031..6242 of consensus"
13917..14217
/note="AluSx repeat: matches 1..303 of consensus"
14218..14516
/note="L1MD2 repeat: matches 5660..6031 of consensus"
14522..14691
/note="L1M4 repeat: matches 2357..2493 of consensus"
14692..15025
/note="AluSx repeat: matches 1..310 of consensus"
15026..15055
/note="L1M4 repeat: matches 2324..2357 of consensus"
15158..15588
/note="L1MB4 repeat: matches 5725..6183 of consensus"
15605..15784
/note="AluDb repeat: matches 1..271 of consensus"
15785..15850
/note="L1MB4 repeat: matches 5667..5733 of consensus"
15851..16168
/note="AluSc repeat: matches 1..306 of consensus"
16169..16739
/note="L1MB4 repeat: matches 5176..5667 of consensus"
16740..17052
/note="AluSx repeat: matches 1..312 of consensus"
17053..17519
/note="L1MB4 repeat: matches 4696..5176 of consensus"
17541..19073

[illegible]

| | | | |
|---------------------------|---|---|-----------------------|
| Db | 36995 | TGCTCAAGGGTGGATCTTCTTGACCTTCCACAGCTGGAGTAGTAGGTTCTAAAGTGA | 36936 |
| QY | 856 | TGATGATCGTGTCCATTACCCAAAGCCTTTAAATCCCTCATGCT-CAGTACACGAG | 914 |
| Db | 36935 | TTCACTCCACCATCCCAATCTCCCTAAGCCTTTGGATCCCTTCTCTACATTAATCAAG | 36876 |
| QY | 915 | GCAGTCTAGCATTTTC | 930 |
| Db | 36875 | AGAGATCAGGCATTTTC | 36860 |
| RESULT | 12 | | |
| AF070718 | | | |
| LOCUS | AF070718 | 166832 bp | DNA |
| DEFINITION | Homo sapiens BAC clone 529F11 from 8q21, complete sequence. | | PRI |
| ACCESSION | AF070718 | | |
| NID | 93283634 | | |
| VERSION | AF070718.1 | GI:3283634 | |
| KEYWORDS | | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 166832) | | |
| TITLE | Tauchi, H., Matsuura, S., Isomura, M., Komatsu, K. and Nakamura, Y. | | |
| JOURNAL | Direct Submission | | |
| COMMENT | Submitted (03-JUN-1998) Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan | | |
| | Mapping and sequence information: | | |
| | This chromosome 8 clone was provided by S. Matsuura, H. Tauchi, and K. Komatsu at Department of Radiation Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan. | | |
| | Sequencing was performed by H. Tauchi, M. Isomura and Y. Nakamura at Human Genome Center, Institute of Medical Science, The University of Tokyo. | | |
| FEATURES | | | |
| source | Source information: Clone 529F11 is isolated from the human BAC library available from Research Genetics, Inc. The library contains cloned DNA from the male fibroblast cell line 978SK. Vector: pBelOBAC11. | | |
| | Location/Qualifiers | | |
| | 1..166832 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="8" | | |
| | /map="8q21" | | |
| | /clone="529F11" | | |
| | /sex="male" | | |
| | /cell_line="978SK" | | |
| | /cell_type="fibroblast" | | |
| | /note="selection using chloramphenicol; orientation of the sequence is from centromere to telomere" | | |
| STS | 40651..41020 | | |
| | /note="STS marker D8S1724" | | |
| BASE COUNT | 52344 a 32042 c 31175 g 51271 t | | |
| ORIGIN | | | |
| Query Match | 11.6%; | Score 129.8; | DB 11; Length 166832; |
| Best Local Similarity | 66.2%; | Pred. No. 5.4e-26; | |
| Matches 280; Conservative | 0; | Mismatches 132; | Indels 11; Gaps 6; |
| QY | 518 | GAACACTGTGATCAACTAGCCAGCACCATAGTTCT-CCGAAGTCAGACTATCATGATTAC | 576 |
| Db | 47221 | GAACACCGTGATTAAATAGCAATGCGAGAGCTCTACATGAGTCAGACTATTTAAATGC | 47280 |
| QY | 577 | TGCTTGACTGTGCTGTCTATTGTACTAACCATGCCGATGTTTAGGTGAA-----ATT | 630 |
| Db | 47281 | TGCTTGCCCTCTGCTGTCCATTATGGTAGTACTATGCCACCTTGCCCTTGATGGTTGAGTG | 47340 |

| | | | |
|----|-------|--|-------|
| QY | 631 | GCGTCACTTGCCCTCAACCATCTGGTAATCCAGTTATCCTCACTGAATTGAGATTTCGTG | 690 |
| Dd | 47341 | CCACCACCTTGTCCTGCCACTTTGGGATCCAATTATTCACACTGCATTTAAATTTTGTA | 47400 |
| QY | 691 | CTTCAGTGTACGCCAATCCCAC-ATAATTCTGACCTACAGAGGTGAGGGATCATATAGC | 749 |
| Dd | 47401 | GTTAGTAGACTGTGTTCTCTGCTGTGATCTGGCATACAGAGAAGCAATCACAGAGC | 47460 |
| QY | 750 | TCTTCAAGGATGCTGCTACTCCCCCTACAATAATTCATTCTC-CTGTGTAGTGAAGGTG | 808 |
| Dd | 47461 | TCTTCAAGGAGGCGAGSTGCTGCCCTCACAAATCTATTTCACCAAAGTGTGTGTAAGG GTA | 47520 |
| QY | 809 | CGCCCTCTGAGCCTCCACGAGGTGGGTGTGACAGT-CACAATGATGAATGTATGATCGTG | 867 |
| Dd | 47521 | TGTCCTCTGAGACCTCCACAGCTGGGATGAGTAGTGCTAAAGTGACTAATCCACTCTAGCA | 47580 |
| QY | 868 | TTCCCATTAACCAAGCCTTT-AAATCCCTPATGTCTCAGTACACCAAGGCGAGTCTAGCA | 926 |
| Dd | 47581 | TCCCAATCTCCCTAAGCCTTTGGATTTCTTCCTCTACATTTAAACCAAGGAGATCAGGCA | 47640 |
| QY | 927 | TTT 929 | |
| Dd | 47641 | TTT 47643 | |

| | |
|------------|---|
| RESULT | 13 |
| LOCUS | AC005906/c |
| DEFINITION | AC005906 185952 bp DNA PRI 30-JAN-1999 Homo sapiens 12p13.3 BAC RPLC11-429A20 (Roswell Park Cancer Institute Human BAC Library) complete sequence. |
| ACCESSION | AC005906 |
| NID | g4165009 |
| VERSION | AC005906.1 GI:4165009 |
| KEYWORDS | HTG. |
| SOURCE | human. |
| ORGANISM | Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| REFERENCE | Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H., Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S., Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W., Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A., Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A. Direct Submission |
| TITLE | Unpublished |
| JOURNAL | 2 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (03-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 3 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (30-DEC-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 4 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (20-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | 5 (bases 1 to 185952) |
| REFERENCE | Worley,K.C. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| JOURNAL | |

COMMENT

On Jan 20, 1999 this sequence version replaced gi:4079596.
INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics

| | |
|--|------------|
| Contig length: | 185961 |
| Phrap values in estimate: | 98505 |
| Average error rate (BCM-Phrap estimate): | 5.51369e-0 |
| Fraction of Phrap values less than 40 : | 0.00108624 |
| Number of consensus changing edits: | 4 |
| Number of N's in consensus : | 0 |

| Position | Original+Context | Edited+Context |
|----------|----------------------|----------------------|
| 9492 | aaagtaac(t)tttttttt | aaagtaac(c)tttttttt |
| 36770 | tggtgtggg(n)agcgggag | tggtgtggg(g)agcgggag |
| 107439 | ctggccttg(n)cgtacact | ctggccttg(a)cgtacact |
| 109251 | gtcccggc(n)ccctactac | gtcccggc(c)ccctactac |

| Quality | Position | Bases with BCM-Phrap value < 20 | Surrounding Sequence |
|---------|----------|---------------------------------|----------------------|
| 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 |
| 10 | 10 | 10 | 10 |
| 11 | 11 | 11 | 11 |
| 12 | 12 | 12 | 12 |
| 13 | 13 | 13 | 13 |
| 14 | 14 | 14 | 14 |
| 15 | 15 | 15 | 15 |
| 16 | 16 | 16 | 16 |
| 17 | 17 | 17 | 17 |
| 18 | 18 | 18 | 18 |
| 19 | 19 | 19 | 19 |
| 20 | 20 | 20 | 20 |
| 21 | 21 | 21 | 21 |
| 22 | 22 | 22 | 22 |
| 23 | 23 | 23 | 23 |
| 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 |
| 29 | 29 | 29 | 29 |
| 30 | 30 | 30 | 30 |
| 31 | 31 | 31 | 31 |
| 32 | 32 | 32 | 32 |
| 33 | 33 | 33 | 33 |
| 34 | 34 | 34 | 34 |
| 35 | 35 | 35 | 35 |
| 36 | 36 | 36 | 36 |
| 37 | 37 | 37 | 37 |
| 38 | 38 | 38 | 38 |
| 39 | 39 | 39 | 39 |
| 40 | 40 | 40 | 40 |
| 41 | 41 | 41 | 41 |
| 42 | 42 | 42 | 42 |
| 43 | 43 | 43 | 43 |
| 44 | 44 | 44 | 44 |
| 45 | 45 | 45 | 45 |
| 46 | 46 | 46 | 46 |
| 47 | 47 | 47 | 47 |
| 48 | 48 | 48 | 48 |
| 49 | 49 | 49 | 49 |
| 50 | 50 | 50 | 50 |
| 51 | 51 | 51 | 51 |
| 52 | 52 | 52 | 52 |
| 53 | 53 | 53 | 53 |
| 54 | 54 | 54 | 54 |
| 55 | 55 | 55 | 55 |
| 56 | 56 | 56 | 56 |
| 57 | 57 | 57 | 57 |
| 58 | 58 | 58 | 58 |
| 59 | 59 | 59 | 59 |
| 60 | 60 | 60 | 60 |
| 61 | 61 | 61 | 61 |
| 62 | 62 | 62 | 62 |
| 63 | 63 | 63 | 63 |
| 64 | 64 | 64 | 64 |
| 65 | 65 | 65 | 65 |
| 66 | 66 | 66 | 66 |
| 67 | 67 | 67 | 67 |
| 68 | 68 | 68 | 68 |
| 69 | 69 | 69 | 69 |
| 70 | 70 | 70 | 70 |
| 71 | 71 | 71 | 71 |
| 72 | 72 | 72 | 72 |
| 73 | 73 | 73 | 73 |
| 74 | 74 | 74 | 74 |
| 75 | 75 | 75 | 75 |
| 76 | 76 | 76 | 76 |
| 77 | 77 | 77 | 77 |
| 78 | 78 | 78 | 78 |
| 79 | 79 | 79 | 79 |
| 80 | 80 | 80 | 80 |
| 81 | 81 | 81 | 81 |
| 82 | 82 | 82 | 82 |
| 83 | 83 | 83 | 83 |
| 84 | 84 | 84 | 84 |
| 85 | 85 | 85 | 85 |
| 86 | 86 | 86 | 86 |
| 87 | 87 | 87 | 87 |
| 88 | 88 | 88 | 88 |
| 89 | 89 | 89 | 89 |
| 90 | 90 | 90 | 90 |
| 91 | 91 | 91 | 91 |
| 92 | 92 | 92 | 92 |
| 93 | 93 | 93 | 93 |
| 94 | 94 | 94 | 94 |
| 95 | 95 | 95 | 95 |
| 96 | 96 | 96 | 96 |
| 97 | 97 | 97 | 97 |
| 98 | 98 | 98 | 98 |
| 99 | 99 | 99 | 99 |
| 100 | 100 | 100 | 100 |

----- Distribution of Quality < 40 Bases -----

| | |
|-----|---|
| 100 | * |
| 90 | * |
| 80 | * |
| 70 | * |
| 60 | * |
| 50 | * |
| 40 | * |
| 30 | * |
| 20 | * |

101
01

5 10 15 20 25 30 35 40
Phrap Value Range

Version: 1.01 xfort.

FEATURES Location/Qualifiers

Source

1.185952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC111-429A20"
/chromosome="12p13.3"
2451.2743
/rpt_family="L1P"
2734.2959
/rpt_family="L1PA11"
complement(3054.3118)
/rpt_family="L2"
complement(3332.3426)
/rpt_family="MERS3"
5492.6248
/rpt_family="L1ME2"
6249.6548
/rpt_family="AluY"
6549.6742
/rpt_family="L1ME2"
complement(6743.6825)
/rpt_family="(CA)n"
6847.7184
/rpt_family="L1ME2"
complement(7202.7315)
/rpt_family="AluJo"
7316.7421
/rpt_family="L1MA10"
7951.8224
/rpt_family="L1MC/D"
9327.9418
/rpt_family="L1M1"
complement(9499.9765)
/rpt_family="AluY"
9832.9928
/rpt_family="MSTA"
10248.10368
/rpt_family="MSTA"
10374.10668
/rpt_family="AluSc"
10669.10901
/rpt_family="MSTA"
12777.12960
/rpt_family="MIR"
13055.13459
/rpt_family="L1TR20"
13754.13808
/rpt_family="MIR"
complement(15214.15393)
/rpt_family="AluJb"
15863.16122
/rpt_family="L1PB3"
16203.16361
/rpt_family="MIR"
complement(16372.16554)
/rpt_family="MIR"
complement(16838.16957)
/rpt_family="(CA)n"
complement(22903.23067)
/rpt_family="MERSB"
23484.23972
/rpt_family="MLT2FA"
24144.24174
/rpt_family="(CAAAA)n"

repeat_region 24212..25128
/rpt_family="MERVL"
repeat_region complement(25147..25701)
/rpt_family="MLT2D"
repeat_region 25705..29985
/rpt_family="HERVL"
repeat_region 30191..30523
/rpt_family="MLT2E"

Query Match 11.5%; Score 128.8; DB 11; Length 185952;
Best Local Similarity 68.0%; Pred. No. 1e-25;
Matches 227; Conservative 0; Mismatches 97; Indels 10; Gaps 3;

OY 518 GAACACTGTGATCAACTAGCCAGCACCATAGTCTCCGAAGTCAGACTATCATGATTACT 577
|||||
DB 26093 GAACACTGTGATCAACTAGCTAATGCCATAGGTCTGTGTGAGGCTATTCTGATTCT 26034
OY 578 GTGTTGACTGTGCTGTCTATTGTACTAACCATTGCCGA--TGTTAGGTGAATTAGCGT 634
|||||
DB 26033 GCCTTGACTTGTGTGTCCATTAAAGTAACACACCCCAATTATTGTTAGTGTCTGT 25974
OY 635 CACTTGCCCTCAACCACTTGCTGTATCCAGTATTCCTCACTGAATTGAGATTCTCTGCTTC 694
|||||
DB 25973 CACTTGCCCTCAACCACTTGCTGTATCCAGTATTCCTCACTGAATTGAGATTCTCTGCTTC 25914
OY 695 AGTGTGAGCCATTCCCA-CATAATTCTGACTACAGAGGTGAGGAGCATATAGCTCTT 753
|||||
DB 25913 AGTGTGAGCCATTCCCA-CATAATTCTGACTACAGAGGTGAGGAGCATATAGCTCTT 25854
OY 754 CAAGCATGCTGTGACTCCCTCCACAAATTCATTCTCTCTGTGTAGTGAAGTGGCGCC 813
|||||
DB 25853 CATGATGCTGTGAGCTCCCTCCACAAATTCATTCTCTCTGTGTAGTGAAGTGGCGCC 25800
OY 814 TCTGAGCCTCCAGGCTGGGTGTGACAGTCA 847
|||||
DB 25799 TTTGATACTTCTTGGGTGTGACAGTCTCA 25766

RESULT 14

HS44N10/c 208643 bp DNA HTG 12-NOV-1998
LOCUS Homo sapiens chromosome 12 clone 44N10, WORKING DRAFT SEQUENCE, in
DEFINITION unordered pieces.
ACCESSION 297197
NID 93873505
VERSION 297197.1 GI:3873505
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
TITLE Primates; Catarrhini; Hominiidae; Homo.
JOURNAL 1 (bases 1 to 208643)

COMMENT
Direct Submission
Submitted (12-NOV-1998) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
hunquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 16, 1998 this sequence version replaced gi:2225928.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: 44N10 contig_ID: 03188 length: 28290 bp
Unfinished: 44N10 Contig_ID: 03967 length: 7858 bp Unfinished:
44N10 Contig_ID: 04550 length: 16612 bp Unfinished: 44N10
Contig_ID: 03869 length: 1065 bp Unfinished: 44N10 Contig_ID: 00878
length: 24110 bp Unfinished: 44N10 Contig_ID: 04458 length: 26725
bp Unfinished: 44N10 Contig_ID: 04588 length: 6852 bp Unfinished:
44N10 Contig_ID: 04631 length: 9284 bp Unfinished: 44N10 Contig_ID:
02809 length: 1644 bp Unfinished: 44N10 Contig_ID: 04439 length:
1307 bp Unfinished: 44N10 Contig_ID: 04403 length: 1278 bp

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:57:15 ; Search time 273 Seconds
(without alignments)
1025.513 Million cell updates/sec

Title: US-09-030-606-177
Perfect score: 1119
Sequence: 1 GCGCACTCGCAGCCCTGCGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 1119 | 100.0 | 1119 | 1 | V58648 | Prostate tumour sp |
| 2 | 1119 | 100.0 | 1119 | 1 | V61253 | CDNA sequence of p |
| 3 | 361.4 | 32.3 | 1265 | 1 | V58645 | Prostate tumour sp |
| 4 | 361.4 | 32.3 | 1265 | 1 | V61250 | CDNA sequence of p |
| 5 | 352.2 | 31.5 | 1167 | 1 | V58647 | Prostate tumour sp |
| 6 | 352.2 | 31.5 | 1167 | 1 | V61252 | CDNA sequence of p |
| 7 | 350.4 | 31.3 | 871 | 1 | V37495 | Human prostate-spe |
| 8 | 339.8 | 30.4 | 1386 | 1 | V11855 | Homo sapiens Tub 1 |
| 9 | 338.8 | 30.3 | 1248 | 1 | V58644 | Prostate tumour sp |
| 10 | 338.8 | 30.3 | 1248 | 1 | V61249 | CDNA sequence of p |
| 11 | 338 | 30.2 | 402 | 1 | X41114 | Human secreted pro |
| 12 | 257.4 | 23.0 | 1459 | 1 | V58646 | Prostate tumour sp |
| 13 | 257.4 | 23.0 | 1459 | 1 | V61251 | CDNA sequence of p |
| 14 | 153.8 | 13.7 | 234 | 1 | V58522 | Prostate tumour sp |
| 15 | 153.8 | 13.7 | 234 | 1 | V61168 | CDNA sequence of p |
| 16 | 100.4 | 9.0 | 1476 | 1 | X16295 | Human kallikrein e |
| 17 | 93 | 8.3 | 907 | 1 | N70905 | Human kallikrein-1 |
| 18 | 91 | 8.1 | 558 | 1 | Q20557 | PAK1191 encoding 1 |
| 19 | 91 | 8.1 | 379 | 1 | Q20555 | BamHI-SacI fragmen |
| 20 | 91 | 8.1 | 555 | 1 | Q20556 | PAK191 encoding 11 |
| 21 | 91 | 8.1 | 309 | 1 | Q20558 | PAK1181 encoding 1 |
| 22 | 91 | 8.1 | 309 | 1 | Q20559 | PAK1181 encoding 1 |
| 23 | 89.8 | 8.0 | 5406 | 1 | N93197 | Fragment of clone |
| 24 | 85 | 7.6 | 738 | 1 | N71049 | Sequence encoding |
| 25 | 81 | 7.2 | 986 | 1 | Q81203 | Human stratum corn |
| 26 | 81 | 7.2 | 1089 | 1 | T39783 | Human amyloid prec |
| 27 | 79 | 7.1 | 450 | 1 | Q20560 | PAK1181 encoding 1 |
| 28 | 76.4 | 6.8 | 832 | 1 | T05147 | Prepro-hk2 kallikr |
| 29 | 76.4 | 6.8 | 832 | 1 | V06609 | Prostate-specific |
| 30 | 76.4 | 6.8 | 832 | 1 | V06604 | Prostate-specific |
| 31 | 76.4 | 6.8 | 832 | 1 | V32938 | Wild-type human Ka |
| 32 | 76.4 | 6.8 | 832 | 1 | V70340 | Prostate-specific |
| 33 | 76.4 | 6.8 | 832 | 1 | X08946 | Human Kallikrein 2 |
| 34 | 76.2 | 6.8 | 798 | 1 | T46175 | Prostate-specific |
| 35 | 75.2 | 6.7 | 925 | 1 | N93196 | Monkey recombinant |
| 36 | 73.4 | 6.6 | 6139 | 1 | V70354 | Coding strand of n |
| 37 | 73.2 | 6.5 | 1466 | 1 | V32496 | Prostate specific |
| 38 | 72.8 | 6.5 | 992 | 1 | T91054 | Human prostate spe |
| 39 | 72.6 | 6.5 | 1462 | 1 | T91055 | Human prostate spe |
| 40 | 71.6 | 6.4 | 200 | 1 | V60379 | mRNA target sequen |
| 41 | 71.6 | 6.4 | 766 | 1 | T05149 | Pro-hk2 kallikrein |
| 42 | 71.6 | 6.4 | 766 | 1 | V06603 | Prostate-specific |
| 43 | 71.6 | 6.4 | 766 | 1 | V70341 | Prostate-specific |

ALIGNMENTS

| | | | | | | |
|----|----|-----|------|---|--------|-------------------|
| 44 | 68 | 6.1 | 1729 | 1 | T04864 | Prostate-specific |
| 45 | 68 | 6.1 | 1728 | 1 | T35867 | Prostate-specific |

RESULT 1
V58648
ID V58648 standard; CDNA; 1119 BP.
AC V58648;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
FH Key
FT CDS Location/Qualifiers
FT 34..528
FT /*tag= a
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69389.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 116-117; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 100.0%; Score 1119; DB 1; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCGCACTCGCAGCCCTGCGAGCGCGGCACTGTCATGGAAGAATGTTCTGCTGGGC 60
DB 1 GCGCACTCGCAGCCCTGCGAGCGCGGCACTGTCATGGAAGAATGTTCTGCTGGGC 60
QY 61 GTCTGTGTCATCCGCGAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCTACACC 120
DB 61 GTCTGTGTCATCCGCGAGTGGGTGCTGTCAAGCCGACACTGTTCCAGAACTCTACACC 120
QY 121 ATCGGGCTGGGCTGCGACAGTCTTGAGGCGCGACCAAGAGCGGAGCCAGATGGTGAG 180
DB 121 ATCGGGCTGGGCTGCGACAGTCTTGAGGCGCGACCAAGAGCGGAGCCAGATGGTGAG 180
QY 181 GCCAGCTCTCCGTACAGCGACCCAGAGTACAAAGACCCCTGCTCGCTAACGACCTCATG 240
DB 181 GCCAGCTCTCCGTACAGCGACCCAGAGTACAAAGACCCCTGCTCGCTAACGACCTCATG 240
QY 241 CTCATCAAGTTGAGCAATCCGTGTCGAGTGTGACACACCATCCGAGCATCAGATTGCT 300
DB 241 CTCATCAAGTTGAGCAATCCGTGTCGAGTGTGACACACCATCCGAGCATCAGATTGCT 300
QY 301 TCGCAGTGGCCCTACCGCGGGGAACCTTGTGCTTCTGCTGGGCTCTGCGGAAC 360
DB 301 TCGCAGTGGCCCTACCGCGGGGAACCTTGTGCTTCTGCTGGGCTCTGCGGAAC 360
QY 361 GATGCTGTGATTGCCATCCAGTCCAGACTGTGGGAGGCTGGAGTGTGAGAAGCTTTCC 420

|||||
Db 361 GATGCTGTGATGCCATCCAGTCCAGACTGTGGAGGCTGGAGGTGAGAAAGCTTCC 420
QY 421 CAACCCCTGGCAGGGTTGTACCATTTCCGCACTTCCAGTGCAGAGAGCGTCCCTGCATC 480
Db 421 CAACCCCTGGCAGGGTTGTACCATTTCCGCACTTCCAGTGCAGAGAGCGTCCCTGCATC 480
QY 481 CTCACCTGGTCTACTACTGCTCATGCTCACCAGCACTGTGATCAACTAGCCAG 540
Db 481 CTCACCTGGTCTACTACTGCTCATGCTCACCAGCACTGTGATCAACTAGCCAG 540
QY 541 CACCATAGTCTCCGAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600
Db 541 CACCATAGTCTCCGAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600
QY 601 ACTAACCATGCCGATGTTAGGTGAATTAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
Db 601 ACTAACCATGCCGATGTTAGGTGAATTAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
QY 661 CAGTTATCTCACTGAATTGAGATTCTCTGCTCAGTGTGAGCCATTCCACATAATTTC 720
Db 661 CAGTTATCTCACTGAATTGAGATTCTCTGCTCAGTGTGAGCCATTCCACATAATTTC 720
QY 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCAACAA 780
Db 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCAACAA 780
QY 781 TTCATTTCTCTGTGTAGTGAAGTGGCCCTCTGAGGCTCCAGGGTGGGTGCA 840
Db 781 TTCATTTCTCTGTGTAGTGAAGTGGCCCTCTGAGGCTCCAGGGTGGGTGCA 840
QY 841 GGTCACAATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTTAATCCCTCATG 900
Db 841 GGTCACAATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTTAATCCCTCATG 900
QY 901 CTCAGTACACCAAGGAGGCTAGCATTTCTTCAATTAGTGTGCTGCTCATGCA 960
Db 901 CTCAGTACACCAAGGAGGCTAGCATTTCTTCAATTAGTGTGCTGCTCATGCA 960
QY 961 ACCACCTCAGGACTCTGGATTCTCTGCTAGTGTGAGCTCCGATGCTGCTCTGGG 1020
Db 961 ACCACCTCAGGACTCTGGATTCTCTGCTAGTGTGAGCTCCGATGCTGCTCTGGG 1020
QY 1021 GAGGTGAGGAGAGGCCCATGTTCATTTGGGATCTGTGAGTGTAAACATTAAGTGC 1080
Db 1021 GAGGTGAGGAGAGGCCCATGTTCATTTGGGATCTGTGAGTGTAAACATTAAGTGC 1080
QY 1081 TTAATAACAGAGCTGTGATGTTAAAAA 1119
Db 1081 TTAATAACAGAGCTGTGATGTTAAAAA 1119

RESULT 2

V61253
ID V61253 standard; cdna; 1119 BP.
AC V61253;
DT 06-JAN-1999 (first entry)
DE cdna sequence of prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI: 98-609886/51.
DR P-PSDB; W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
used in a vaccine for the treatment of prostate cancer
PS - Claim 5; Page 108-109; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cdna library obtained by
CC subtracting a prostate tumour cdna expression library with a normal
CC tissue cdna library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 100.0%; Score 1119; DB 1; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCACTCCGACGCCCTGGCAGCGCGCACTGTGATGAAAAAGCAATTGTTCTGCTGGGC 60
Db 1 GCGCACTCCGACGCCCTGGCAGCGCGCACTGTGATGAAAAAGCAATTGTTCTGCTGGGC 60
QY 61 GTCCTGTGATCCGAGTGGGTGCTGTGAGCCGCACTGTTCCAGACTCTACACC 120
Db 61 GTCCTGTGATCCGAGTGGGTGCTGTGAGCCGCACTGTTCCAGACTCTACACC 120
QY 121 ATCGGCTGGCCCTGCACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGTGAG 180
Db 121 ATCGGCTGGCCCTGCACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGTGAG 180
QY 181 GCCAGCTCTCCGTACGGCACCAGAGTACAACAGACCCTTGCTCGTAACGACCTCATG 240
Db 181 GCCAGCTCTCCGTACGGCACCAGAGTACAACAGACCCTTGCTCGTAACGACCTCATG 240
QY 241 CTCATCAAGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGGAGCATGACATTGCT 300
Db 241 CTCATCAAGTTGGACGAATCCGTGTCCGAGTGTGACACCATCCGGAGCATGACATTGCT 300
QY 301 TCGCAGTGGCCCTACCGCGGGGAACTCTTGCCCTGTTCTGCTGGGTCTGCTGGCAAC 360
Db 301 TCGCAGTGGCCCTACCGCGGGGAACTCTTGCCCTGTTCTGCTGGGTCTGCTGGCAAC 360
QY 361 GATGCTGTGATGTCATTCAGATCCAGACTGTGGAGGCTGGAGTGTGAAGCTTTCC 420
Db 361 GATGCTGTGATGTCATTCAGATCCAGACTGTGGAGGCTGGAGTGTGAAGCTTTCC 420
QY 421 CAACCCCTGGCAGGGTTGTACCAATTCGGCACTTCCAGTGCAGAGAGCGTCCCTGCATC 480
Db 421 CAACCCCTGGCAGGGTTGTACCAATTCGGCACTTCCAGTGCAGAGAGCGTCCCTGCATC 480
QY 481 CTCACCTGGTCTACTACTGCTCATGCTCACCAGCACTGTGATCAACTAGCCAG 540
Db 481 CTCACCTGGTCTACTACTGCTCATGCTCACCAGCACTGTGATCAACTAGCCAG 540
QY 541 CACCATAGTCTCCGAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600
Db 541 CACCATAGTCTCCGAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600
QY 601 ACTAACCATGCCGATGTTAGGTGAATTAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
Db 601 ACTAACCATGCCGATGTTAGGTGAATTAGCGTCACTTGGCCTCAACCATCTTGGTATC 660
QY 661 CAGTTATCTCACTGAATTGAGATTCTCTGCTCAGTGTGAGCCATTCCACATAATTTC 720
Db 661 CAGTTATCTCACTGAATTGAGATTCTCTGCTCAGTGTGAGCCATTCCACATAATTTC 720
QY 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCAACAA 780
Db 721 TGACCTACAGAGGTGAGGATCATATAGCTCTTCAAGAGTGTGACTCCCTCAACAA 780
QY 781 TTCATTTCTCTGTGTAGTGAAGTGGCCCTCTGAGGCTCCAGGGTGGGTGCA 840
Db 781 TTCATTTCTCTGTGTAGTGAAGTGGCCCTCTGAGGCTCCAGGGTGGGTGCA 840
QY 841 GGTCACAATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTTAATCCCTCATG 900
Db 841 GGTCACAATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTTAATCCCTCATG 900


```

QY 901 CTGAGTACACCAGGCGAGGCTAGCATTTCTTCATTTAGTGTATGCTGTCCATTCATGCA 960
Db 901 CTCAGTACACCAGGCGAGGCTAGCATTTCTTCATTTAGTGTATGCTGTCCATTCATGCA 960
QY 961 ACCACCTCAGGACTCCTGATTCCTGCTAGTTGAGCTCCTGCATGCTGCTCCTTGGG 1020
Db 961 ACCACCTCAGGACTCCTGATTCCTGCTAGTTGAGCTCCTGCATGCTGCTCCTTGGG 1020
QY 1021 GAGGTGAGGAGAGAGGGCCCATGTTTCAATGGGATCTGTGCAAGTTGTACACATTAGTGC 1080
Db 1021 GAGGTGAGGAGAGAGGGCCCATGTTTCAATGGGATCTGTGCAAGTTGTACACATTAGTGC 1080
QY 1081 TTAATAACAGAGAAGCTGTGATGTTAAAAA 1119
Db 1081 TTAATAACAGAGAAGCTGTGATGTTAAAAA 1119

RESULT 3
V58645
ID V58645 standard; cDNA; 1265 BP.
AC V58645:
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KW Prostate tumour specific gene; human; prostate cancer; detection;
therapy; ss.
OS Homo sapiens.
PN W09837418-A2.
PF 27-AUG-1998.
PD 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR MPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.3%; Score 361.4; DB 1; Length 1265;
Best Local Similarity 99.7%; Pred. No. 6.5e-98;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAACGAATTTCTGCTCGGGCG 61
Db 8 CGCACTCGCAGCCCTGGCAGGCGGCACTGTCATGGAACGAATTTCTGCTCGGGCG 67
QY 62 TCCTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACACTGTTCCAGAACTCTACACCA 121
Db 68 TCCTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACACTGTTCCAGAACTCTACACCA 127
QY 122 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGAGG 181
Db 128 TCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGCTGAGG 187
QY 182 CCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCCTTGCTCGCTAACGACCTCATGC 241
Db 188 CCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCCTTGCTCGCTAACGACCTCATGC 247
QY 242 TCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTT 301
Db 248 TCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTT 307

```

| | | | |
|---|---|---|-----|
| OY | 302 | CGCAGTGCCTACCCGCGGGGAACCTTGCCCTCGTTTCTGGCTGGGGTCTGCTGCGCAACG | 361 |
| Db | 308 | CGCAGTGCCTACCCGCGGGGAACCTTGCCCTCGTTTCTGGCTGGGGTCTGCTGCGCAACG | 367 |
| OY | 362 | ATG 364 | |
| Db | 368 | GTG 370 | |
| RESULT 4 | | | |
| ID | V61250 | standard; cDNA; 1265 BP. | |
| AC | V61250; | | |
| DT | 06-JAN-1999 | (first entry) | |
| DE | cDNA sequence of prostate tumour clone P703 splice variant DE2. | | |
| KW | Prostate; cancer; tumour; vaccine; immunogen; clone; ss. | | |
| OS | Homo sapiens. | | |
| PN | WO9837093-A2. | | |
| PD | 27-AUG-1998. | | |
| PF | 25-FEB-1998; U03492. | | |
| PR | 09-FEB-1998; US-020956. | | |
| PR | 25-FEB-1997; US-806099. | | |
| PR | 01-AUG-1997; US-904804. | | |
| PA | (CORI-) CORIXA CORP. | | |
| PI | Dillon DC, Xu J; | | |
| PT | WPI: 98-609886/51. | | |
| DR | Polypeptides comprising immunogenic portions of prostate proteins - | | |
| PS | used in a vaccine for the treatment of prostate cancer | | |
| CS | Claim 3; Page 105-106; 130pp; English. | | |
| CC | The present sequence is a new DNA which encodes an immunogenic portion | | |
| CC | of a prostate tumour protein. The encoded immunogen, or the DNA itself, | | |
| CC | can be used as a vaccine for the treatment of prostate cancer. The DNA | | |
| CC | was identified by analysis of a subtracted cDNA library obtained by | | |
| CC | subtracting a prostate tumour cDNA expression library with a normal | | |
| CC | tissue cDNA library. | | |
| SQ | Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T; | | |
| Query Match 32.3%; Score 361.4; DB 1; Length 1265; Best Local Similarity 99.7%; Pred. No. 6.5e+98; Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | |
| OY | 2 | CGCACTCGCAGCCCTGGCAGGCGGCCACTGTCATGGAACAAGATTGTTCTGCTCGGCG | 61 |
| Db | 8 | CGCACTCGCAGCCCTGGCAGGCGGCCACTGTCATGGAACAAGATTGTTCTGCTCGGCG | 67 |
| OY | 62 | TCTGTGTCATCCGACGTGGGTGCTGTACAGCCGACACTGTTCCAGAAGCTCTACACA | 121 |
| Db | 68 | TCTGTGTCATCCGACGTGGGTGCTGTACAGCCGACACTGTTCCAGAAGCTCTACACA | 127 |
| OY | 122 | TGGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAGAGCCAGGGAGCCAGATGGTGAGG | 181 |
| Db | 128 | TGGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAGAGCCAGGGAGCCAGATGGTGAGG | 187 |
| OY | 182 | CCAGCCTCTCCGTACGCGCACCCAGAGTAGACAACAGACCCTTGCTCGCTAACGACCTCATGC | 241 |
| Db | 188 | CCAGCCTCTCCGTACGCGCACCCAGAGTAGACAACAGACCCTTGCTCGCTAACGACCTCATGC | 247 |
| OY | 242 | TCATCAAGTTGAGCAAATCCGTGTCGAGTGTGACACATCCGAGCATCAGCATGCTT | 301 |
| Db | 248 | TCATCAAGTTGAGCAAATCCGTGTCGAGTGTGACACATCCGAGCATCAGCATGCTT | 307 |
| OY | 302 | CGCAGTGCCTACCCGCGGGGAACCTTGCCCTCGTTTCTGGCTGGGGTCTGCTGCGCAACG | 361 |
| Db | 308 | CGCAGTGCCTACCCGCGGGGAACCTTGCCCTCGTTTCTGGCTGGGGTCTGCTGCGCAACG | 367 |
| OY | 362 | ATG 364 | |
| Db | 368 | GTG 370 | |

V58647
ID V58647 standard; cDNA; 1167 BP.
AC V58647;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KW Prostate tumour specific gene; human; prostate cancer; detection;
therapy; ss.
OS Homo sapiens.
FH Key
FT CDS
Location/Qualifiers
28.645
/*tag= a
PN MO9837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 31.5%; Score 352.2; DB 1; Length 1167;
Best Local Similarity 96.3%; Pred. No. 3.4e-95;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTCGGCGCTCTGG 67
DB 2 CGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTCGGCGCTCTGG 61
QY 68 TGCATCCGCACTGGTGTCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATCGGGC 127
DB 62 TGCATCCGCACTGGTGTCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATCGGGC 121
QY 128 TGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCC 187
DB 122 TGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCC 181
QY 188 TCTCCGTACGGCAGCCAGAGTACAACAGACCCTGCTCGCTAACGACCTCATGCTCATCA 247
DB 182 TCTCCGTACGGCAGCCAGAGTACAACAGACTCTGCTCGCTAACGACCTCATGCTCATCA 241
QY 248 AGTTGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATTCGCTCGCAGT 307
DB 242 AGTTGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATTCGCTCGCAGT 301
QY 308 GCCCTACCGCGGGGAACCTCTGCTGCTTCTGCTGGGCTGCTGCGGCAACGATGCTG 367
DB 302 GCCCTACCGCGGGGAACCTCTGCTGCTTCTGCTGGGCTGCTGCGGCAACGAGAA 361
QY 368 TGATTGCCATCCAG 381
DB 362 TGCTTACCGTGTGCTG 375

RESULT 6
ID V61252 standard; cDNA; 1167 BP.
AC V61252;
FT

DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 31.5%; Score 352.2; DB 1; Length 1167;
Best Local Similarity 96.3%; Pred. No. 3.4e-95;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTCGGCGCTCTGG 67
DB 2 CGCAGCCCTGGCAGCGCGCACTGTCATGGAAGAATGTTCTGCTCGGCGCTCTGG 61
QY 68 TGCATCCGCACTGGTGTCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATCGGGC 127
DB 62 TGCATCCGCACTGGTGTCTGTCAGCCGACACACTGTTCCAGAACTCCTACACCATCGGGC 121
QY 128 TGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCC 187
DB 122 TGGGCTGACAGTCTTGAGGCCGACCAAGAGCCAGAGCCAGATGGTGAGGCCAGCC 181
QY 188 TCTCCGTACGGCAGCCAGAGTACAACAGACCCTGCTCGCTAACGACCTCATGCTCATCA 247
DB 182 TCTCCGTACGGCAGCCAGAGTACAACAGACTCTGCTCGCTAACGACCTCATGCTCATCA 241
QY 248 AGTTGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATTCGCTCGCAGT 307
DB 242 AGTTGACGAATCCGTGTCCAGTCTGACACCATCCGAGCATCAGCATTCGCTCGCAGT 301
QY 308 GCCCTACCGCGGGGAACCTCTGCTGCTTCTGCTGGGCTGCTGCGGCAACGATGCTG 367
DB 302 GCCCTACCGCGGGGAACCTCTGCTGCTTCTGCTGGGCTGCTGCGGCAACGAGAA 361
QY 368 TGATTGCCATCCAG 381
DB 362 TGCTTACCGTGTGCTG 375

RESULT 7
ID V37495 standard; DNA; 871 BP.
AC V37495;
DT 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
OS Homo sapiens.
FH Key
FT CDS
Location/Qualifiers
31.777
/*tag= a
/transl_except= (pos:367..369, aa:Xaa)
/transl_except= (pos:412..414, aa:Xaa)
FT


```
FT      /transl_except- (pos:424. .426, aa:Xaa)
FT      /product= "HPSK protein"
FT      /note= "Xaa = unknown"
PN      WO9820117-A1.
PD      14-MAY-1998.
PF      31-OCT-1997; U20051.
PR      05-NOV-1996; US-744026.
PA      (INCY-) INCYTE PHARM INC.
PI      Bandman O, Goll SR;
DR      WPI; 98-286933/25.
DR      P-PSDB; W60592.
PT      New isolated prostate-specific kallikrein - used to develop products
PT      for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT      hyperplasia
PS      Claim 5; Fig 1A-C; 68bp; English.
CC      This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC      containing an expression vector comprising the HPSK nucleic acid sequence
CC      can be used to produce the protein recombinantly. The HPSK products can
CC      be used for the diagnosis of conditions or diseases associated with
CC      expression of HPSK such as prostate carcinoma and benign prostate
CC      hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC      modulate its activity can be used for the preparation of treatment of
CC      such conditions or diseases. The products can also be used for detection
CC      and drug screening, especially for the detection of prostate-specific
CC      kallikrein (PSK).
SQ      Sequence      871 BP;      166 A;      260 C;      258 G;      184 T;
```

Query Match 31.3%; Score 350.4; DB 1; Length 871;
Best Local Similarity 95.0%; Pred. No. 1e-94;
Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```
OY      3 GCACCTCGACCCCTGGCAGGCGGCGACATGTCATGAAACGAATTGTTCTGCTGGCGGT 62
DB      129 GCATTCACACCCCTGGCAGGCGGCGACATGTCATGAAACGAATTGTTCTGCTGGCGGT 188
OY      63 CCTGTGCATCCGACAGTGGTGTGTCAGCCGCACACTGTTCCAGAACTCCTACACCAT 122
DB      189 CCTGTGCATCCGACAGTGGTGTGTCAGCCGCACACTGTTCCAGAACTCCTACACCAT 248
OY      123 CGGGCTGGGCTGCACAGTCTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGC 182
DB      249 CGGGCTGGGCTGCACAGTCTGAGGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGC 308
OY      183 CAGCCTCTCCGACAGGACCCAGAGTACAAGACCCTTGTGCTGCTAACGACCTCATGCT 242
DB      309 CAGCCTCTCCGACAGGACCCAGAGTACAAGACCCTTGTGCTGCTAACGACCTCATGNT 368
OY      243 CATCAAGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCATCAGCATTCGCTTC 302
DB      369 CATCAAGTTGAGCAATCCGTGTCGAGTCTGACAAACATCCGAGNATCAGCATTCGCTTC 428
OY      303 GCAGTGGCCCTACCGCGGGAACCTTTCCTGCTTCTGCTGGGCTCTGCTGGCAACGA 362
DB      429 GCAGTGGCCCTACCGCGGGAACCTTTCCTGCTTCTGCTGGGCTCTGCTGGCAACGG 488
OY      363 TGCTGTGATGCCATCCAG 381
DB      489 CAGAATGCCCTACCGTCTG 507
```

RESULT 8

```
V11855
ID      V11855 standard; cDNA; 1386 BP.
AC      V11855;
DE      Homo sapiens Tub Interactor (hri-1) gene.
KW      serine protease; tub interactor; treatment; obesity; cachexia;
KW      anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW      neurodegenerative disease; Alzheimer's disease; drug screening;
KW      Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW      amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
OS      Homo sapiens.
```

```
FT      Key
FT      CDS
FT      Location/Qualifiers
FT      2..701
FT      /*tag= a
FT      /product= hri-1 protein
FT      /note= "putative serine protease"
```

```
PN      WO9812302-A1.
PD      26-MAR-1998.
PF      05-SEP-1997; U15627.
PR      21-JUL-1997; US-897340.
PR      17-SEP-1996; US-715032.
PA      (MILL-) MILENMIUM PHARM INC.
PI      Errada PR, Gimeno CJ;
DR      WPI; 98-217246/19.
DR      P-PSDB; W59129.
PT      Tub interactor genes - used to develop products for the treatment
PT      of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT      diabetes
PS      Claim 10; Fig 1; 120pp; English.
CC      The sequence is that of the Tub Interactor gene hri-1 which
CC      codes for a putative serine protease. TI genes function
CC      in biochemical pathways involved in weight control and
CC      related disorders. The products can be used for treating
CC      weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC      or a related disorder such as diabetes. The products can
CC      also be used to modulate cell cycle progression and apoptosis.
CC      They can be used for treating neurodegenerative diseases
CC      which are characterised by apoptosis, including Alzheimer's
CC      disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC      lateral sclerosis or spinocerebellar degenerations. The
CC      products can also be used for detection, diagnosis and
CC      drug screening.
SQ      Sequence      1386 BP;      318 A;      490 C;      321 G;      249 T;
```

Query Match 30.4%; Score 339.8; DB 1; Length 1386;
Best Local Similarity 93.9%; Pred. No. 1.8e-91;
Matches 369; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

```
OY      1 GCGCAGTCGACGCCCTGGCAGGCGGCGACATGTCATGGAACGAATTGTTCTGCTCGGGC 60
DB      14 GCGCAGTCGACGCCCTGGCAGGCGGCGACATGTCATGGAACGAATTGTTCTGCTCGGGC 73
OY      61 GTCTGTGTCATCCGACAGTGGTGTGTCAGCCGCACACTGTTCCAGAA----- 111
DB      74 GTCTGTGTCATCCGACAGTGGTGTGTCAGCCGCACACTGTTCCAGAAAGTAGTCAG 133
OY      111 --CTCTACACCATCGGGCTGGGCTGTCACAGTCTGAGGCCGACCAAGAGCCAGGAGC 168
DB      134 AGCTCTACACCATCGGGCTGGGCTGTCACAGTCTGAGGCCGACCAAGAGCCAGGAGC 193
OY      169 CAGATGTGGAGGCCAGCCTTCCTGACGGCACCCAGAGTACAACAGACCCCTTCTGCTGCT 228
DB      194 CAGATGTGGAGGCCAGCCTTCCTGACGGCACCCAGAGTACAACAGACCCCTTCTGCTGCT 253
OY      229 AACGACCTCATGCTCATCAAGTTGAGCAATCCGTCGAGTCTGACACCATCCGAGC 288
DB      254 AACGACCTCATGCTCATCAAGTTGAGCAATCCGTCGAGTCTGACACCATCCGAGC 313
OY      289 ATCAGATGCTCTTGCAGTGCCTACCGCGGGAACCTTTCCTGCTTTCGCTGGGGT 348
DB      314 ATCAGATGCTCTTGCAGTGCCTACCGCGGGAACCTTTCCTGCTTTCGCTGGGGT 373
OY      349 CTGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
DB      374 CTGCTGGCGAAGCGACGAATGCCCTACCGTCTG 406
```

RESULT 9

```
V58644
ID      V58644 standard; cDNA; 1248 BP.
AC      V58644;
DE      Homo sapiens Prostate specific gene clone DE1.
OS      Homo sapiens.
```


KW Prostate tumour specific gene; human; prostate cancer; detection;
KM therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 217..696
FT /*tag= a
PD WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69387.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 112; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
Best Local Similarity 93.9%; Pred. No. 3.4e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCAGTCGACGCCCTGGCAGCGCGCACTGTCATGAAACGAATGTTCTGCTCGGCG 61
DB 35 CGCAGTCGACGCCCTGGCAGCGCGCACTGTCATGAAACGAATGTTCTGCTCGGCG 94
QY 62 TCCTGGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTCCAGAA----- 111
DB 95 TCCTGGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTCCAGAAAGTAGTCAGA 154
QY 111 -CTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAAGAGCGAGGCC 169
DB 155 GCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAAGAGCGAGGCC 214
QY 170 AGATGGTGAAGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAGACCCCTGCTGCTA 229
DB 215 AGATGGTGAAGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAGACCCCTGCTGCTA 274
QY 230 ACGACCTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACACCATCCGAGCA 289
DB 275 ACGACCTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACACCATCCGAGCA 334
QY 290 TCAGCATGCTTTCGAGTGCCTTACCGCGGGGGAATCTTGCTTCTGCTGGGCTC 349
DB 335 TCAGCATGCTTTCGAGTGCCTTACCGCGGGGGAATCTTGCTTCTGCTGGGCTC 394
QY 350 TGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
DB 395 TGCTGGCGAAGCATGCTGTGATGCCATCCAG 426

RESULT 10
V61249
ID V61249 standard; cDNA; 1248 BP.
AC V61249;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant D61.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837418-A2.

PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
Best Local Similarity 93.9%; Pred. No. 3.4e-91;
Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCAGTCGACGCCCTGGCAGCGCGCACTGTCATGAAACGAATGTTCTGCTCGGCG 61
DB 35 CGCAGTCGACGCCCTGGCAGCGCGCACTGTCATGAAACGAATGTTCTGCTCGGCG 94
QY 62 TCCTGGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTCCAGAA----- 111
DB 95 TCCTGGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTCCAGAAAGTAGTCAGA 154
QY 111 -CTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAAGAGCGAGGCC 169
DB 155 GCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGACCAAGAGCGAGGCC 214
QY 170 AGATGGTGAAGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAGACCCCTGCTGCTA 229
DB 215 AGATGGTGAAGCCAGCCTCTCCGTACGGCAGCCAGAGTACAAGACCCCTGCTGCTA 274
QY 230 ACGACCTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACACCATCCGAGCA 289
DB 275 ACGACCTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACACCATCCGAGCA 334
QY 290 TCAGCATGCTTTCGAGTGCCTTACCGCGGGGGAATCTTGCTTCTGCTGGGCTC 349
DB 335 TCAGCATGCTTTCGAGTGCCTTACCGCGGGGGAATCTTGCTTCTGCTGGGCTC 394
QY 350 TGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
DB 395 TGCTGGCGAAGCATGCTGTGATGCCATCCAG 426

RESULT 11
X41114
ID X41114 standard; cDNA; 402 BP.
AC X41114;
DT 17-JUN-1999 (first entry)
DE Human secreted protein 5' EST SEQ ID NO:58.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
KW Homo sapiens.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 99-153778/13.
DR P-PSDB; Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 30.2%; Score 338; DB 1; Length 402;
Best Local Similarity 96.8%; Pred. No. 3.4e-91;
Matches 360; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 2 CGCACTCGCAGCCCTGGCAGCGGCGCACTGTCATGGAACGAATTGTCGTCGGCG 61
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
23 CGCACTCGCAGCCCTGGCAGCGGCGCACTGTCATGGAACGAATTGTCGTCGGCG 82
QY 62 TCCTGGTGCATCCGAGTGGTGTGTGACGCCGACACTGTTCCAGAA----- 111
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
83 TCCTGGTGCATCCGAGTGGTGTGTGACGCCGACACTGTTCCAGAAAGTGCAGAC 142
QY 111 -CTCTACACCATCGGGCTGGGCTGTGACAGTCTTGAGGCGCCAGCAAGCCAGGAGCC 169
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
143 GCTCTACACCATCGGGCTGGGCTGTGACAGTCTTGAGGCGCCAGCAAGCCAGGAGCC 202
QY 170 AGATGGTGGAGGCGGCTCTCCGTACGCGACCCAGAGTACACAGAGCCCTTGCTGCTA 229
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
203 AGATGGTGGAGGCGGCTCTCCGTACGCGACCCAGAGTACACAGAGCCCTTGCTGCTA 262
QY 230 ACGACCTCATGTCATCAAGTTGGACGAATCCGTGCGAGTCTGACACCATCCGGAGCA 289
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
263 ACGACCTCATGTCATCAAGTTGGACGAATCCGTGCGAGTCTGACACCATCCGGAGCA 322
QY 290 TCAGCATTTGCTTCGAGTGGCCCTACCGGGGGAATCTTGCTGCTTGCTGGGCTC 349
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
323 TCAGCATTTGCTTCGAGTGGCCCTACCGGGGGAATCTTGCTGCTTGCTGGGCTC 382
QY 350 TGCTGGCGAAGC 361
Db ||||||||||||||||
383 TGCTGGCGAAGC 394

RESULT 12
V58646
ID V58646 standard; cDNA; 1459 BP.
AC V58646;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KM therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;
Best Local Similarity 97.8%; Pred. No. 6.3e-67;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 98 ACTGTTCCAGAACTCCTACACCATCGGGCTGGCCTGCACAGTCTTGAGCGCCAGCAAG 157
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
25 AGTGAGTGCAGAGCTCCTACACCATCGGGCTGGCCTGCACAGTCTTGAGCGCCAGCAAG 84
QY 158 AGCCAGGAGGACGAGATGTGAGGCGCAGCCTCTCCGTACGCGCACCCAGAGTACACAGAC 217
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
85 AGCCAGGAGGACGAGATGTGAGGCGCAGCCTCTCCGTACGCGCACCCAGAGTACACAGAC 144
QY 218 CTTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAAGATCCGTGTCGAGTGTGACA 277
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
145 CTTGCTCGCTAACGACCTCATGCTCATCAAGTTGAGCAAGATCCGTGTCGAGTGTGACA 204
QY 278 CCATCCGGAGCATGACATGCTTCCGAGTGGCCCTACCGGGGGAATCTTGCTGCTTT 337
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
205 CCATCCGGAGCATGACATGCTTCCGAGTGGCCCTACCGGGGGAATCTTGCTGCTTT 264
QY 338 CTGGCTGGGCTGCTGCGAAGCATG 364
Db ||||||||||||||||||||||||||||
265 CTGGCTGGGCTGCTGCGAAGCATG 291

RESULT 13
V61251
ID V61251 standard; cDNA; 1459 BP.
AC V61251;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;
Best Local Similarity 97.8%; Pred. No. 6.3e-67;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 98 ACTGTTCCAGAACTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGCCGACCAAG 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 AGTGAGTGCAGAGCTCCTACACCATCGGGCTGGCCCTGCACAGTCTTGAGCCGACCAAG 84
QY 158 AGCCAGGAGCCAGATGTGTGAGGCCACCCCTCTCCGTACGCGACCCAGAGTACACAGAC 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 AGCCAGGAGCCAGATGTGTGAGGCCACCCCTCTCCGTACGCGACCCAGAGTACACAGAC 144
QY 218 CCTGCTCGCTAACGACCTCATGCTCAAGTTGGACGAATCCGTGTCGAGTGTGACA 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 CCTGCTCGCTAACGACCTCATGCTCAAGTTGGACGAATCCGTGTCGAGTGTGACA 204
QY 278 CCATCCGAGCATCAGCATGCTTCGAGTGCCCTACCGCGGGAACCTCTGCGCTGTT 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 CCATCCGAGCATCAGCATGCTTCGAGTGCCCTACCGCGGGAACCTCTGCGCTGTT 264
QY 338 CTGGCTGGGCTCTGCTGGCGAAGCATG 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 CTGGCTGGGCTCTGCTGGCGAAGCATG 291
```

RESULT 14

```
V58522
ID V58522 standard; cDNA; 234 BP.
AC V58522;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone P20.
KW Prostate tumour specific gene; human; prostate cancer; detection;
    therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
    for detecting and treating prostate cancers
PS Claim 1; Page 56; 141p; English.
CC This sequence represents a human prostate tumour specific gene, and can
    be used in the method of the invention. The method is for detecting
    prostate cancer comprises contacting a biological sample with an agent
    able to bind an immunogenic portion of a prostate protein (such as
    encoded by this sequence). An antibody which binds to an immunogenic
    portion of the prostate protein, and the method can be used to detect,
    monitor progression of, or treat prostate cancers. The antibody may
    also be conjugated to a therapeutic agent for use in therapy of prostate
    cancers.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;
```

Query Match 13.7%; Score 153.8; DB 1; Length 234;
Best Local Similarity 93.1%; Pred. No. 1.8e-36;
Matches 161; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 209 ACAACAGACCCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ACAACAGACCCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 60
QY 269 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCCTACCGGGGAAGTCTT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCCTACCGGGGAAGTCTT 120
QY 329 GCCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATGCCATCCAG 173
```

RESULT 15

```
V61168
ID V61168 standard; cDNA; 234 BP.
AC V61168;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P20.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
    used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 53-54; 130p; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
    of a prostate tumour protein. The encoded immunogen, or the DNA itself,
    can be used as a vaccine for the treatment of prostate cancer. The DNA
    was identified by analysis of a subtracted cDNA library obtained by
    CC subtracting a prostate tumour cDNA expression library with a normal
    CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;
```

Query Match 13.7%; Score 153.8; DB 1; Length 234;
Best Local Similarity 93.1%; Pred. No. 1.8e-36;
Matches 161; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 209 ACAACAGACCCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ACAACAGACCCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGCCG 60
QY 269 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCCTACCGGGGAAGTCTT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 AGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGTGCCCTACCGGGGAAGTCTT 120
QY 329 GCCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATGCCATCCAG 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCCTCGTTTCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATGCCATCCAG 173
```

Search completed: September 25, 1999, 09:57:21
Job time: 8092 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:30:09 ; Search time 165.01 Seconds
(without alignments)
621.696 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119

Sequence: 1 GCGCACATCGCAGCCCTGGCA.....ATGTTAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NDC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/lna/5C_COMB.seq:*

4: /cgn2_6/ptodata/2/lna/5D_COMB.seq:*

5: /cgn2_6/ptodata/2/lna/PCTUS9_COMB.seq:*

6: /cgn2_6/ptodata/2/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 350.4 | 31.3 | 871 | 3 | US-08-744-026-2 |
| 2 | 81 | 7.2 | 986 | 4 | US-08-557-146-1 |
| 3 | 81 | 7.2 | 1089 | 5 | PCT-US96-04294-1 |
| 4 | 81 | 7.2 | 1089 | 5 | PCT-US96-04294-3 |
| 5 | 76.4 | 6.8 | 832 | 5 | PCT-US95-06157-5 |
| 6 | 72.8 | 6.5 | 992 | 2 | US-08-358-782D-13 |
| 7 | 72.6 | 6.5 | 1462 | 2 | US-08-358-782D-14 |
| 8 | 72 | 6.4 | 200 | 3 | US-08-931-981A-1 |
| 9 | 71.6 | 6.4 | 766 | 5 | PCT-US95-06157-9 |
| 10 | 68 | 6.1 | 1729 | 4 | US-08-844-024-1 |
| 11 | 63.6 | 5.7 | 760 | 5 | PCT-US95-06157-7 |
| 12 | 62.4 | 5.6 | 833 | 4 | US-08-790-137-2 |
| 13 | 58.4 | 5.2 | 2259 | 4 | US-08-845-998-3 |
| 14 | 56.8 | 5.1 | 2218 | 4 | US-08-845-998-5 |
| 15 | 50.8 | 4.5 | 957 | 3 | US-08-684-862-11 |
| 16 | 49.6 | 4.4 | 1454 | 3 | US-08-467-155A-2 |
| 17 | 49.6 | 4.4 | 1454 | 3 | US-08-628-198-2 |
| 18 | 49.6 | 4.4 | 1454 | 5 | PCT-US96-07343-2 |
| 19 | 46 | 4.1 | 840 | 3 | US-08-684-862-12 |
| 20 | 45.6 | 4.1 | 732 | 3 | US-08-361-395-2 |
| 21 | 45 | 4.0 | 111 | 3 | US-08-454-720A-42 |
| 22 | 40.8 | 3.6 | 925 | 1 | US-07-990-301A-3 |
| 23 | 40.4 | 3.6 | 7218 | 2 | US-08-232-463-14 |
| 24 | 40.2 | 3.6 | 734 | 3 | US-08-650-129-1 |
| 25 | 40.2 | 3.6 | 821 | 3 | US-08-650-129-2 |
| 26 | 40.2 | 3.6 | 866 | 3 | US-08-650-129-3 |
| 27 | 38.4 | 3.4 | 1096 | 3 | US-08-684-862-8 |
| 28 | 36.4 | 3.3 | 1554 | 3 | US-08-469-486-1 |
| 29 | 36 | 3.2 | 970 | 1 | US-08-148-910-3 |
| 30 | 36 | 3.2 | 2033 | 1 | US-08-148-910-14 |
| 31 | 36 | 3.2 | 970 | 2 | US-08-448-937A-3 |
| 32 | 36 | 3.2 | 2033 | 2 | US-08-448-937A-14 |
| 33 | 35.8 | 3.2 | 959 | 3 | US-08-568-031-1 |
| 34 | 35.8 | 3.2 | 959 | 4 | US-08-966-319-1 |
| 35 | 35.4 | 3.2 | 2743 | 1 | US-08-396-479B-3 |
| 36 | 35.4 | 3.2 | 2743 | 2 | US-08-818-823-3 |
| 37 | 35.4 | 3.2 | 2749 | 4 | US-08-124-981A-1 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 38 | 35.4 | 3.2 | 2853 | 5 | PCT-US94-07297-36 | Sequence 36, Appl |
| 39 | 35 | 3.1 | 807 | 2 | US-08-270-584A-1 | Sequence 1, Appl |
| 40 | 35 | 3.1 | 807 | 4 | US-08-765-192-1 | Sequence 1, Appl |
| 41 | 34.8 | 3.1 | 867 | 1 | US-07-990-301A-1 | Sequence 1, Appl |
| 42 | 34.2 | 3.1 | 1333 | 3 | US-08-684-862-9 | Sequence 9, Appl |
| 43 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-1 | Sequence 1, Appl |
| 44 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-3 | Sequence 3, Appl |
| 45 | 33.8 | 3.0 | 681 | 1 | US-07-929-198-5 | Sequence 5, Appl |

ALIGNMENTS

RESULT 1

US-08-744-026-2

Sequence 2, Application US/08744026

Patent No. 5786148

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goll, Surya K.

TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC

TITLE OF INVENTION: KALLIKREIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/744,026

FILING DATE: Herewith

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0154 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 871 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY:

CLONE: Consensus

US-08-744-026-2

Query Match 31.3%; Score 350.4; DB 3; Length 871;
Best Local Similarity 95.0%; Pred. No. 5.6e-103;
Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| OY | 3 | GCACTGCGAGCCCTGGCAGCGGCGCATGTCATGGAACGAATTGTTCTGCTGGCGCT | 62 |
| Db | 129 | GCAATCCAGCCCTGGCAGCGGCGCATGTCATGGAACGAATTGTTCTGCTGGCGCT | 188 |
| OY | 63 | CCTGTGATCCGAGTGCGTGTGTCAGCCGACACACTGTTCCAGAACTCCTACACAT | 122 |
| Db | 189 | CCTGTGATCCGAGTGCGTGTGTCAGCCGACACACTGTTCCAGAACTCCTACACAT | 248 |

[illegible]

```

RESULT 2
US-08-557-146-1
: Sequence 1, Application US/08557146
: Patent No. 5834290
: GENERAL INFORMATION:
: APPLICANT: Egellrud, Torbjorn
: APPLICANT: Hansson, Lennart
: TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
: TITLE OF INVENTION: Enzyme (SCCE)
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: White & Case, Patent Department
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2787
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/557,146
: FILING DATE: 14-DEC-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Sterner, Richard J.
: REGISTRATION NUMBER: 35,372
: REFERENCE/DOCKET NUMBER: 1103326-181
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 819-8783
: TELEFAX: (212) 354-8113
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 986 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 25..786
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 25..90
: FEATURE:
: NAME/KEY: mat_peptide

```

LOCATION: 112..783
US-08-557-146-1

| | | | | |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match | 7.2%; | Score 81; | DB 4; | Length 986; |
| Best Local Similarity | 53.7%; | Pred. No. 1.8e-16; | | |
| Matches 196; | Conservative 0; | Mismatches 160; | Indels 9; | Gaps 1; |

| | | | |
|----|-----|---|-----|
| QY | 6 | CTGGAGCCCCCTGGSCAGGCGGCACCTGGTTCATGGAAACGAATTGTCTCTGCTCGGGCCGTCCT | 65 |
| Db | 141 | CTCCACCACCATGGCAGGTGGCCCTGCTCAGTGCCAATCAGCTCCACTGCCGAGGCGCTCT | 200 |
| QY | 66 | GGTGCATCCCGCAGTGGGTGCTGTCAAGCCGCACACTGTTTCCAGAACA CTCTACACCATCGG | 125 |
| Db | 201 | GGTCAATGAGCGCTGGGTGCTCTCACTGCGGCCCACTGCAAGATGAATGAGTACACCGTGCA | 260 |
| QY | 126 | GCTGGGCGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAG | 185 |
| Db | 261 | CCTGGGCAGTGATACGCTGGGCGACAGAGAG - - - - - CTCAGAGGATCAAGCGCTTC | 311 |
| QY | 186 | CCTCTCCGTACGCGCACCCAGATACACAGACCCTTGCTGCTAAGCACCCTCATGCTCAT | 245 |
| Db | 312 | GAAGTCAATCCGCGCACCCCGGCTACTCCACACAGACCACATTAATGACCTCATGCTCGT | 371 |
| QY | 246 | CAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCATTGTCTGCA | 305 |
| Db | 372 | GAAAGCTCAATAGCCAGCGCAGCGCTGTATCCATGTGTGAAGAAAGTCAGGCTGCCCTCCG | 431 |
| QY | 306 | GTGCCCTACCGCGGGGGAAC TCTTGCTCTGTTCTGGCTGGGGTCTGCTGGCGAACGATGC | 365 |
| Db | 432 | CTGCGAACCCTTGGAACCAACCTGTACTGTCTCCGGCTGGGGCACTACCAAGAGCCCA GA | 491 |
| QY | 366 | TGTGA 370 | |
| Db | 492 | TGTGA 496 | |

RESULT 3
PCT-US96-04294-1/c
Sequence 1, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
TITLE OF INVENTION: RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bialock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1089 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US96-04294-1

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTCGCAGCCCTGGCAGCGGCGACATGTCATGAAACGAATTGTTCTGCTCGGCGCTCT 65
DB 857 CTCACACCATGCGAGGTGGCCCTGCTCAGTGCATCAGCTCAGCTCGGAGGCGCTCT 798
QY 66 GGTGCATCCGCAAGTGGTGTCTGTCAGCCGCAACACTGTTCCAGACTCCTACACCATCGG 125
DB 797 GGTCAATGAGCGCTGGTGTCTCAGTCCGCGCCACTGCAAGATGATGATACACCGTGA 738
QY 126 GCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGGAGATGTGAGGCCAG 185
DB 737 CTTGGGCGAGTGTATCGCTGGGCGACAGAGAG-----CTCAGAGGATCAAGGCCCTC 687
QY 186 CCTCTCCGTCAGCGCAACCCAGAGTACACAGAACCCCTGCTCGCTAACGACCTCATGCTCAT 245
DB 686 GAAGTCATTCGCGCCACCCCGGCTACTCTCACACAGAACCCATGTTAATGACCTCATGCTCT 627
QY 246 CAAGTTGACGAATCCGTGTCCGAGTGTACACCATCCGAGCATCAGCATGCTTCGCA 305
DB 626 GAAGCTCAATAGCCAGGCGAGCTGTATCATGCTGTAAGAAAGTACAGGCTGCCCTCCG 567
QY 306 GTGCCCTACCGCGGGAACCTCTGCTCTGCTTCTGCTGGGCTGTGCTGCGAAGCATGC 365
DB 566 CTGCGAACCCTTGAACCACTGTACTGTCTCCGCTGGGCGACTACACGAGGCCAGAGA 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 4

PCT-US96-04294-3/c
Sequence 3, Application PC/TUS9604294

GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:

NAME: Blalock, Donna R.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1089 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
PCT-US96-04294-3

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTCGCAGCCCTGGCAGCGGCGACATGTCATGAAACGAATTGTTCTGCTCGGCGCTCT 65
DB 857 CTCACACCATGCGAGGTGGCCCTGCTCAGTGCATCAGCTCAGCTCGGAGGCGCTCT 798
QY 66 GGTGCATCCGCAAGTGGTGTCTGTCAGCCGCAACACTGTTCCAGACTCCTACACCATCGG 125
DB 797 GGTCAATGAGCGCTGGTGTCTCAGTCCGCGCCACTGCAAGATGATGATACACCGTGA 738
QY 126 GCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGGAGATGTGAGGCCAG 185
DB 737 CTTGGGCGAGTGTATCGCTGGGCGACAGAGAG-----CTCAGAGGATCAAGGCCCTC 687
QY 186 CCTCTCCGTCAGCGCAACCCAGAGTACACAGAACCCCTGCTCGCTAACGACCTCATGCTCAT 245
DB 686 GAAGTCATTCGCGCCACCCCGGCTACTCTCACACAGAACCATGTTAATGACCTCATGCTCT 627
QY 246 CAAGTTGACGAATCCGTGTCCGAGTGTACACCATCCGAGCATCAGCATGCTTCGCA 305
DB 626 GAAGCTCAATAGCCAGGCGAGCTGTATCATGCTGTAAGAAAGTACAGGCTGCCCTCCG 567
QY 306 GTGCCCTACCGCGGGAACCTCTGCTCTGCTTCTGCTGGGCTGTGCTGCGAAGCATGC 365
DB 566 CTGCGAACCCTTGAACCACTGTACTGTCTCCGCTGGGCGACTACACGAGGCCAGAGA 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 5

PCT-US95-06157-5
Sequence 5, Application PC/TUS9506157

GENERAL INFORMATION:
APPLICANT: Mayo Foundation for Medical Education
APPLICANT: and Research
APPLICANT: Hybritech Incorporated
APPLICANT: Tindall, Donald J.
APPLICANT: Young, Charles Y.F.
APPLICANT: Saedi, Mohammed S.
TITLE OF INVENTION: Recombinant HK2 Polypeptide
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
STREET: 3500 IDS Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25


```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06157
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Raasch, Kevin W.
: REGISTRATION NUMBER: 35,561
: REFERENCE/DOCKET NUMBER: 150.148W01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: TELEFAX: 612-339-3061
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 10..792
PCT-US95-06157-5

```

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 6.8%; | Score 76.4; | DB 5; | Length 832; |
| Best Local Similarity | 63.0%; | Pred. No. 4.8e-15; | | |
| Matches 136; Conservative | 0; | Mismatches 76; | Indels 4; | Gaps 1; |

| | | | | | | | |
|----|-----|--------------|-----------|-------------------|------------------------|---------------|-----|
| QY | 311 | CTACCGGGGGA | CTCTTGCC | CTGTTTCTGCGCTGGGG | CTCTGCTGGCGA | ACGATGCTGTGA | 370 |
| | | | | | | | |
| Db | 5 | CCAGCATGTGGG | GA | CTGGTTCTCTCC | ATGCGCTTGTCTGTGGGG | GTGC | 64 |
| | | | | | | | |
| QY | 371 | TTGCCATCCAGT | CCAGAC | TA | CTGTGGAGGCTGGGAGTGTGAG | AAGCTTTCCCA | 430 |
| | | | | | | | |
| Db | 65 | CCCTCATCCAGT | CTCGGAT | TGTGGG | AGGCTGGGAGTGTGAG | AAGCATTTCCCA | 124 |
| | | | | | | | |
| QY | 431 | AGG--- | GTTGTACCA | ATTTCG | CAACTTCCAGTGC | AAGACGTCTCTGC | 486 |
| | | | | | | | |
| Db | 125 | AGGTGGCTGTGT | ACAGTCA | TGATGGATGGG | CACTCTGGGGGTGTCTGTG | CACCCCACT | 184 |
| | | | | | | | |
| QY | 487 | GGGTGCTCACT | ACTACTGCT | CACTGCAT | CACCCGGA | ACA | 522 |
| | | | | | | | |
| Db | 185 | GGGTGCTCAC | AGCTGCCCA | TTGCTTA | AAGA | GAATA | 220 |
| | | | | | | | |

RESULT 6
 US-08-358-782D-13
 ; Sequence 13, Application US/08358782D
 ; Patent No. 5674682
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Croce, Carlo
 ; APPLICANT: Gomella, Leonard
 ; APPLICANT: Mulholland, S. Grant
 ; APPLICANT: Moreno, Jose
 ; APPLICANT: Fischer, Rainer
 ; TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/358,782D
 ; FILING DATE: 15-DEC-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ;

```

1 NAME: Beardell, Lori Y.
2 REGISTRATION NUMBER: 34,293
3 REFERENCE/DOCKET NUMBER: TUV-1327
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: 215-568-3100
6 TELEFAX: 215-568-3439
7 INFORMATION FOR SEQ ID NO: 13:
8 SEQUENCE CHARACTERISTICS:
9 LENGTH: 992 base pairs
10 TYPE: nucleic acid
11 STRANDEDNESS: single
12 TOPOLOGY: linear
13 MOLECULE TYPE: DNA (genomic)
14 HYPOTHETICAL: NO
15 ANTI-SENSE: NO
16
17 US-08-358-782D-13

```

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------|
| Query Match | 6.5%; | Score 72.8; | DB 2; | Length 992; |
| Best Local Similarity | 59.9%; | Pred. No. 7.7e-14; | | |
| Matches 139; | Conservative | 0; | Mismatches 89; | Indels 4; |
| | | | | Gaps 1; |

[illegible]

RESULT 7
US-08-358-782D-14
Sequence 14, Application US/08358782D
Patent No. 5674682
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Gomella, Leonard
APPLICANT: Mulholland, S. Grant
APPLICANT: Moreno, Jose
APPLICANT: Fischer, Rainer
TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,782D
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: TUV-1327
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439

Db 9 CATCCAGTCTCGATGTGTGGAGGCTGGAGTGTGAGAAGCATTTCCCAACCTGGCAGT 68
QY 434 ---GTGTACCATTTGCGCACTTCCAGTGCAGAGACGTCTGCTGCATCCTCACTGGGT 490
Db 69 GGCTGTGTACAGTCATGTGAGGCGACACTGTGGGGGTCTCTGTGTGCACCCCACTGGGT 128

QY 491 GCTCACTACTGCTCACTGCATCACCAGGACA 522
Db 129 GCTCACAGCTGCCCATTTGCCCTAAGAAGATA 160

RESULT 10

US-08-844-024-1

; Sequence 1, Application US/08844024

; Patent No. 5840494

; GENERAL INFORMATION:

; APPLICANT: Katz, Aaron E., et al.

; TITLE OF INVENTION: A Method For Molecular Staging Of

; TITLE OF INVENTION: Prostate Cancer

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/844,024

; FILING DATE: 18-APR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/229,391

; FILING DATE: 15-APR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 43677/JPW/TEP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 977-9550

; TELEFAX: (212) 664-0525

; TELETYPE: 422523 COOP UI

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1729 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 378..1088

; US-08-844-024-1

Query Match 6.1%; Score 68; DB 4; Length 1729;

Best Local Similarity 66.1%; Pred. No. 3.7e-12;

Matches 115; Conservative 0; Mismatches 55; Indels 4; Gaps 1;

QY 361 GATGCTGTGATTCGCATCCAGTCCAGACTGTGGAGCTGGAGTGTGAGAAGCTTCC 420
Db 351 GGTGCTGCACCCCTCATCTCTCTCGATTGTGGAGCTGGAGTGCAGAACATTC 410

QY 421 CAACCTGCGCAGG---GTTGACCATTTTCGCAACTTCCAGTGCAGAGACGTCCTGCTG 476
Db 411 CAACCTGCGCAGGCTGTTGAGCTCTCTGTCGACGCGCAGTCTGCGGCGTGTCTGTG 470

QY 477 CATCTCACTGGGTGCTCACTACTGCTCATGTCATCACCAGGACAAGTGTATC 530
Db 471 CACCCAGTGGGTCTCACAGCTACCACTGCATCAGAACAAAAGCGTATC 524

RESULT 11

PCT-US95-06157-7

; Sequence 7, Application PC/TUS9506157

; GENERAL INFORMATION:

; APPLICANT: Mayo Foundation for Medical Education

; APPLICANT: and Research

; APPLICANT: Hybritech Incorporated

; APPLICANT: Tindall, Donald J.

; APPLICANT: Young, Charles Y.F.

; APPLICANT: Saeed, Mohammed S.

; TITLE OF INVENTION: Recombinant HK2 Polypeptide

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schweigman, Lundberg & Woessner, P.A.

; STREET: 3500 IDS Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/06157

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Raasch, Kevin W.

; REGISTRATION NUMBER: 35,561

; REFERENCE/DOCKET NUMBER: 150.148W01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-339-0331

; TELEFAX: 612-339-3061

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 760 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 7..720

; PCT-US95-06157-7

Query Match 5.7%; Score 63.6; DB 5; Length 760;

Best Local Similarity 58.6%; Pred. No. 5.9e-11;

Matches 130; Conservative 0; Mismatches 89; Indels 3; Gaps 1;

QY 3 GCATCCGACAGCCCTGGCAGGCGGCACTGGTCATGGAAGCAATGTCTGCTGGGCGT 62
Db 36 GCATTCACCAACCTGGCAGGTGGCTGTGTACAGTCATGATGGGCACACTGTGGGGTGT 95

QY 63 CCGGTGTCATCCGCACTGGGTGCTGTGACGCGCACACTGTTTCCAGAACTCCTACACCAT 122
Db 96 CCGGTGTCACCCCTGAGGTGCTGTGACAGCTGCCCCATTGCTTAAGAAATAGCCAGGT 155

QY 123 CCGGTGCGCCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGTGAGGC 182
Db 156 CCGGTGCGCCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGTGAGGC 212

QY 183 CAGCCTCTCCGTACGCGACCCAGAGTACACAGACCCCTGCT 224
Db 213 CAGCAGACAGCTTCCACACCCGCTCTACAAATATGAGCCTTCT 254

| | | | |
|----|-----|---|-----|
| QY | 283 | CGGAGCATCAGCAT-----TGCTTCGAGTGCCTTACCGCGGGGAACCTCTTGCCCTGTT | 336 |
| | | | |
| Db | 419 | CACACGTCACGCTGCCCCCTGCTCGAGACCTTCCCCCGGGGATGCCGTGTGGTTC | 478 |
| | | | |
| QY | 337 | TCTGGCTGGGGTCTGCTGGCGAAGATGCTGTGATTGCCATCCAGTCCAGACTGTGGGA | 396 |
| | | | |
| Db | 479 | ACTGGCTGGGGCGACGTGGACATATATGTGTGGTGTGGGACACCGGGAGGCCGGGCCA | 538 |
| | | | |
| QY | 397 | GGCTGGGA | 404 |
| | | | |
| Db | 539 | GGTGGCA | 546 |

```

RESULT 14
US-08-845-998-5
Sequence 5, Application US/08845998
Patent No. 5879892
GENERAL INFORMATION:
APPLICANT: Van Baren, Nicolas
APPLICANT: Coulie, Pierre G.
APPLICANT: De Smet, Charles
APPLICANT: Lucas, Sophie
APPLICANT: Boon, Thierry
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845, 998
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2218 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: NVB352/3
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-08-845-998-5

```

[illegible]

QY 105 --CCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTGAGGCCGACCAGAAGCCA 162
| | | | | | | | | | | | | | |
Db 239 CCGAGCATCAAGATCTGGCCGCCCTCAGGGGTGCAACTGGGGAGCAGCACCTCTACTAC 298

QY 163 GGGAGCCAGATGTTGAGGCCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCTTG 222
| | | | | | | | | | | | | | |
Db 299 CAGGACCAGCTGTCGGGTACGAGCATCATCGTGACCACAGTTCTACATCATCCAG 358

QY 223 CTCGCTAACGAACCTCATGCTCATCAAGTTGGACGAATCCGTTGCCAGTCTGACACCATC 282
| | | | | | | | | | | | | | |
Db 359 ACCGGGGCAGCATCGCCCTGCTGAGCTGGAGAGAGCCCGTAACATCTCCAGCCACATC 418

QY 283 CGGAGCATCAGCAT-----TGCTTCGAGTGCCTACCGGGGGAACTCTTGCTGCTT 336
| | | | | | | | | | | | | | |
Db 419 CACACGTCACAGCTGCCCCCTGCTCGAGACCTTCCCCCGGGGATGCCGTGCTGGTTC 478

QY 337 TCTGGCTGGGGTCTGCTGGGGAACGATGCTGTGATTGCCATCCAGTCCAGACTGTGGGA 396
| | | | | | | | | | | | | | |
Db 479 ACTGGCTGGGGCCACGTTGACATAATATGTTGGGTGTTGGGACAGCGGGAGCGCGGCCA 538

QY 397 GGCTGGGA 404
| | | | |
Db 539 GGTGGGCA 546

```

RESULT 15
US-08-684-862-11
: Sequence 11, Application US/08684862
: Patent No. 5759541
: GENERAL INFORMATION:
: APPLICANT: Bach, Alfred
: APPLICANT: Hillel, Heinz
: APPLICANT: Bialojan, Siegfried
: TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
: TITLE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Keil & Weinlauf
: STREET: 1101 Connecticut Avenue
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM AT-compatible, 80286 processor
: OPERATING SYSTEM: MS-DOS version 5.0
: SOFTWARE: Wordperfect version 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/684,862
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/361,705
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/966,040
: FILING DATE: 30-DEC-1992
: APPLICATION NUMBER: PCT/EP91/01361
: FILING DATE: 19-JUL-1991
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 957 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Agkistrodon rhodostoma
: FEATURE:
: LOCATION: 210 to 911
: OTHER INFORMATION: the coding region shown in (2)(1x)(B)
: OTHER INFORMATION: codes for the protein of SEQ ID NO: 6

```


US-08-684-862-11

Query Match 4.58; Score 50.8; DB 3; Length 957;
Best Local Similarity 48.08; Pred. No. 8.7e-07;
Matches 145; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

```
QY 46 TTGTTCTGCTGGGGCGTCTGTCATCCGAGTGGGTGCTGACGCCACACTGTTTC 105
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 TTGTCGCGCTGGGACTTTGATCCACCAAGATGGGTGCTGCTGCACACTGCGAG 338
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 CAGAACTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGG 165
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 AGTACGATCTCAAGATGAAGTTTGATGATAGCAAAAAGTACAAATGAGGATGAG 398
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 AGCCAGATGCTGAGAGCCAGCCCTCTCCGTACGGCACCCAGAGTACACAGACCCTTGCTC 225
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 CAGACAAGAAACGCAAGGAAAAAGTCAITTTGTCCCAATAAGAAAAACGATGAAGTACTG 458
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGG 285
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 459 GACAAGGACATTATGTGATCAAGCTGAACCATCTGTTAGCAATAGTGAACACATCGCG 518
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 AGCATCAGATTGCTTTCGACAGTGCCTACCGCGGGGAACCTTTCCTGCTGCTGG 345
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 CCTCTCAGCTTGCCCTTCCAGCCCTCCAGTGTGGGCTCATTTGCCATATTATGGGATGG 578
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 GG 347
    ||
DB 579 GG 580
```

Search completed: September 25, 1999, 07:30:10
Job time: 3447 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:28 ; Search time 1811.29 Seconds
(without alignments)
1218.615 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119

Sequence: 1 GCGCACTCGCAGCCCTGCA.....ATGTTAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 361 | 32.3 | 777 | 48 | AI557281 | AI557281 PT2.1_15- |
| 2 | 223.6 | 20.0 | 722 | 48 | AI557025 | AI557025 PT2.1_10- |
| 3 | 214.2 | 19.1 | 415 | 35 | AA551449 | AA551449 nj55e05.s |
| 4 | 205.8 | 18.4 | 569 | 50 | AI686689 | AI686689 tu35g11.x |
| 5 | 170 | 15.2 | 216 | 32 | AA336074 | AA336074 EST40886 |
| 6 | 99.6 | 8.9 | 479 | 46 | AI436093 | AI436093 th92b12.x |
| 7 | 99.6 | 8.9 | 475 | 46 | AI452907 | AI452907 t146g04.x |
| 8 | 92 | 8.2 | 245 | 31 | AA295609 | AA295609 EST100780 |
| 9 | 91.6 | 8.2 | 264 | 32 | AA366980 | AA366980 EST77998 |
| 10 | 82.2 | 7.3 | 429 | 35 | AA594076 | AA594076 nn31a12.s |
| 11 | 81.4 | 7.2 | 407 | 38 | AA776018 | AA776018 ae83g11.s |
| 12 | 81 | 7.2 | 523 | 35 | AA542994 | AA542994 n155b01.s |
| 13 | 76.2 | 6.8 | 486 | 35 | AA535837 | AA535837 nj79g05.s |
| 14 | 76.2 | 6.8 | 572 | 35 | AA579026 | AA579026 nf34f02.s |
| 15 | 74.8 | 6.7 | 253 | 20 | T29510 | T29510 EST82448 Hu |
| 16 | 74 | 6.6 | 546 | 33 | AA388907 | AA388907 mp16b03.r |
| 17 | 72.6 | 6.5 | 336 | 34 | AA506459 | AA506459 nh46a08.s |
| 18 | 72.6 | 6.5 | 548 | 35 | AA578976 | AA578976 nf26g03.s |
| 19 | 72.6 | 6.5 | 315 | 35 | AA593245 | AA593245 nn07e10.s |
| 20 | 72.6 | 6.5 | 902 | 47 | AI525832 | AI525832 PT1.3_06- |
| 21 | 72.6 | 6.5 | 732 | 48 | AI547309 | AI547309 PN001_AH- |
| 22 | 71.6 | 6.4 | 470 | 47 | AI525128 | AI525128 promrna-7 |
| 23 | 71 | 6.3 | 610 | 48 | AI557591 | AI557591 pt2.1-1.F |
| 24 | 70 | 6.3 | 371 | 34 | AA503943 | AA503943 nh38d05.s |
| 25 | 70 | 6.3 | 484 | 36 | AA603529 | AA603529 np15b08.s |
| 26 | 70 | 6.3 | 523 | 36 | AA639901 | AA639901 np08f03.s |
| 27 | 70 | 6.3 | 737 | 48 | AI547285 | AI547285 PN001_AH- |
| 28 | 70 | 6.3 | 454 | 50 | AI685510 | AI685510 tu36b10.x |
| 29 | 69.8 | 6.2 | 703 | 48 | AI546857 | AI546857 PN2.1_07- |
| 30 | 69.4 | 6.2 | 457 | 45 | AI385433 | AI385433 m185b05.y |
| 31 | 69.4 | 6.2 | 595 | 48 | AI557389 | AI557389 PT2.1_6_G |
| 32 | 69 | 6.2 | 619 | 26 | W58737 | W58737 zd23e07.r1 |
| 33 | 69 | 6.2 | 435 | 29 | AA177005 | AA177005 nc03b02.s |
| 34 | 68.8 | 6.1 | 432 | 38 | AA778667 | AA778667 af87e12.s |
| 35 | 67.8 | 6.1 | 340 | 27 | AA038537 | AA038537 m185b05.r |
| 36 | 67.2 | 6.0 | 519 | 42 | AI098337 | AI098337 vg84b03.r |
| 37 | 67.2 | 6.0 | 760 | 42 | AI127299 | AI127299 qb74f07.x |
| 38 | 66.8 | 6.0 | 453 | 34 | AA533652 | AA533652 nj72g08.s |
| 39 | 65.6 | 5.9 | 616 | 48 | AI546898 | AI546898 PN2.1_09- |
| 40 | 65.4 | 5.8 | 457 | 42 | AI078417 | AI078417 oz05e02.x |
| 41 | 65 | 5.8 | 624 | 47 | AI525001 | AI525001 promrna-4 |
| 42 | 64.2 | 5.7 | 397 | 35 | AA573604 | AA573604 nf42d12.s |
| 43 | 64 | 5.7 | 472 | 28 | AA062294 | AA062294 ml70a06.r |
| 44 | 64 | 5.7 | 482 | 28 | AA064644 | AA064644 ml73c10.r |
| 45 | 64 | 5.7 | 568 | 28 | AA068804 | AA068804 rnm62g08.r |

ALIGNMENTS

RESULT 1
LOCUS AI557281 777 bp mRNA
DEFINITION PT2.1_15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557281
NID g4489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 777)
Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source location/Qualifiers

1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 32.3%; Score 361; DB 48; Length 777;
Best Local Similarity 100.0%; Pred. No. 6.9e-96;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGCACTCGACGCCCTGGCAGCGGCACACTGTCATGGAAGAAGATGTTCTGCTCGGGC 60
Db 14 GCGCACTCGACGCCCTGGCAGCGGCACACTGTCATGGAAGAAGATGTTCTGCTCGGGC 73
OY 61 GTCTGTGTCATCCGACAGTGGGTGCTGTCAGCCCGCACACTGTTCCAGAACTCCTACACC 120
Db 74 GTCTGTGTCATCCGACAGTGGGTGCTGTCAGCCCGCACACTGTTCCAGAACTCCTACACC 133
OY 121 ATCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAG 180
Db 134 ATCGGGCTGGGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGGTGAG 193
OY 181 GCCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTCGCTAACGACCTCATG 240
Db 194 GCCAGCTCTCCGTACGCGACCCAGAGTACAACAGACCCCTGCTCGCTAACGACCTCATG 253
OY 241 CTCATCAAGTTGACGAAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCT 300
Db 254 CTCATCAAGTTGACGAAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCT 313
OY 301 TCGCAGTCCCCCTACCGCGGGGAACTCTGCGTCTTGCGTGCGGTCTGCTGGCGAAG 360
Db 314 TCGCAGTCCCCCTACCGCGGGGAACTCTGCGTCTTGCGTGCGGTCTGCTGGCGAAG 373
OY 361 G 361
Db 374 G 374

RESULT 2
LOCUS AI557025 722 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557025
NID g4489388
VERSION AI557025.1 GI:4489388

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 722)
Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source location/Qualifiers

1..722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 146 a 156 c 195 g 176 t 49 others
ORIGIN

Query Match 20.0%; Score 223.6; DB 48; Length 722;
Best Local Similarity 95.6%; Pred. No. 2.1e-55;
Matches 237; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

OY 109 AACTCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 168
Db 115 AGCTTCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGC 174
OY 169 CAGATGTTGAGAGCCACCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCTCGCT 228
Db 175 CAGATGTTGAGAGCCACCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCTCGCT 234
OY 229 AACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCTGACACCATCCGGAGC 288
Db 235 AACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCTGACACCATCCGGAGC 294
OY 289 ATCAGCATGCTTCGAGTGCCTTACCGCGGGGAACTCTTGCC-TGTTTCTGGCTGGG 347
Db 295 ATCAGCATGCTTCGAGTGCCTTACCGCGGGGAACTCTTNCCTTCTGCTGGG 354
OY 348 TCTGCTGG 355
Db 355 TCTGCTTG 362

RESULT 3
LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
DEFINITION nj55e05.s1 NCI_CGAP_P9 Homo sapiens cDNA clone IMAGE:986416 similar to SW:K1KA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR ;, mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 415) |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) |
| JOURNAL | |
| COMMENT | On May 9, 1995 this sequence version replaced gi:802243. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuvaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html

Insert Length: 640 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .415. |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_11b="NCI_CGAP_Pr9"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."

```

| | | | | |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match | 19.18; | Score 214.2; | DB 35; | Length 415; |
| Best Local Similarity | 98.68; | Pred. No. 1e-52; | | |
| Matches 216; Conservative | 0; | Mismatches 3; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 143 | TTGAGGCCCGACCAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGGCACC | 202 |
| Dd | 2 | TCGACGCCCGACCAAGAGCCAGGGAGCCAGACGGTGAGGGCCAGCCTCTCCGTACGGGCACC | 61 |
| QY | 203 | CAGAGTACAACAGAACCCCTTGCTCGCTAAAGAACCTCATGCTCATCAAGTTGGAGGAATCCG | 262 |
| Dd | 62 | CAGAGTACAACAGAACCCCTTGCTCGCTAAAGAACCTCATGCTCATCAAGTTGGAGGAATCCG | 121 |
| QY | 263 | TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCGCAATGCCCTAACC GGCGGA | 322 |
| Dd | 122 | TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCGCAATGCCCTAACC GGCGGA | 181 |
| QY | 323 | ACTCTTGCCCTCGTTTTCTGGCTGGGGGCTGTGCTGGCGAAGC | 361 |
| Dd | 182 | ACTCTTGCCCTCGTTTTCTGGCTGGGGGCTGTGCTGGCGAAGC | 220 |

| RESULT | 4 |
|------------|---|
| AI686689/c | |
| LOCUS | |
| DEFINITION | |
| AI686689 | 569 bp mRNA |
| tu35g11.x1 | NCI_CGAP_Pt28 Homo sapiens cDNA clone IMAGE:2253092 3 |
| similar to | TR:Q92046 Q92046 PREPROTRYPsin PRECURSOR ; mRNA |
| sequence. | |

| | |
|-----------|--------------|
| ACCESSION | AI686689 |
| NID | g4897983 |
| VERSION | AI686689.1 |
| KEYWORDS | GI:4897983 |
| SOURCE | EST. |
| ORGANISM | human. |
| | Homo sapiens |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 569)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNT at:
www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 444.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .569 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

```

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 125 a | 150 c | 170 g | 124 t |
| ORIGIN | | | | |

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------|
| Query Match | 18.4%; | Score 205.8; | DB 50; | Length 569; |
| Best Local Similarity | 94.7%; | Pred. No. 3.4e-50; | | |
| Matches. 213; | Conservative | 0; | Mismatches 12; | Indels 0; |
| | | | Gaps | 0; |

| | | | |
|----|-----|---|-----|
| QY | 157 | GAGCCAGGGAGCCAGATGGTGGAGGCCACGCTCTCCGTACGGCACCAGAGTACAACAGA | 216 |
| | | | |
| Dd | 569 | GAGCCAGGGAGCCAGATGGTGGAGGCCACGCTCTCCGTACGGCACCAGAGTACAACAGA | 510 |
| QY | 217 | CCCTTGCTCGCTAACGACCTCATGCTCAATGAAGTAATCCGTGTCCGAGTCTGAC | 276 |
| | | | |
| Dd | 509 | CCCTTGCTCGCTAACGACCTCATGCTCAATGAAGTAATCCGTGTCCGAGTCTGAC | 450 |
| QY | 277 | ACCATCCGAGACATCAGCATTTGCTTCGAGTGCCTACCGGGGSACTCTTGCCCTGTT | 336 |
| | | | |
| Dd | 449 | ACCATCCGAGACATCAGCATTTGCTTCGAGTGCCTACCGGGGSACTCTTGCCCTGTT | 390 |
| QY | 337 | TCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATTGCCATTCAG | 381 |
| | | | |
| Dd | 389 | TCTGGCTGGGGTCTGCTGGCGAAGCATGCTGTGATTGCCATTCAG | 345 |

| | | | |
|------------|----------|--|-------------|
| RESULT | 5 | | |
| AA336074 | | | |
| LOCUS | | | |
| DEFINITION | AA336074 | 216 bp | mRNA |
| ACCESSION | EST40886 | Endometrial tumor Homo sapiens CDNA 5' | 21-APR-1997 |
| NID | AA336074 | similar to kallikrein family, mRNA sequence. | |
| | 91988560 | | |

VERSION AA336074.1 GI:1988560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 216)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrle,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280
JOURNAL
MEDLINE
COMMENT On Apr 14, 1993 this sequence version replaced gi:692773.
On Apr 14, 1993 this sequence version replaced gi:692773.
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 301869056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
FEATURES
source
1. 216
/organism="Homo sapiens"
/db_xref="ATCC (ihost):138014"
/db_xref="taxon:9606"
/clone_lib="Endometrial tumor"
/sex="female"
/dev_stage="adult"
/note="Organ: endometrium; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 45 a 65 c 63 g 38 t 5 others
ORIGIN
Query Match 15.2%; Score 170; DB 32; Length 216;
Best Local Similarity 97.1%; Pred. No. 8.1e-40;
Matches 170; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 CGCACTCGCAGCCCTGGCAGGGCGGCACTGTCATGGAAGCAATGTTCTGCTCGGGCG 61
DB 42 CGCACTCGCAGCCCTGGCAGGGCGGCACTGTCATGGAAGCAATGTTCTGCTCGGGCG 101
QY 62 TCCTGTGTCATCCGACGTGGGTGCTGTACGGCGACACTGTTCCAGAACTCCTACACCA 121
DB 102 TCCTGTGTCATCCGACGTGGGTGCTGTACGGCGACACTGTTCCAGAACTCCTACACCA 161
QY 122 TCGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGTT 176
DB 162 TCGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGTT 216
RESULT 6

AI436093/c
LOCUS AI436093 479 bp mRNA
DEFINITION th92b12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2126111 3' similar to TR:002710 002710 GAG POLYPEPTIDE. ;
mRNA sequence.
ACCESSION AI436093
NID 94307772
VERSION AI436093.1 GI:4307772
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 479)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT On May 18, 1998 this sequence version replaced gi:3137463.
On May 18, 1998 this sequence version replaced gi:3137463.
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 677 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
FEATURES
source
1. 479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2126111"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NbHSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 143 a 85 c 127 g 124 t
ORIGIN
Query Match 8.9%; Score 99.6; DB 46; Length 479;
Best Local Similarity 62.5%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 134; Indels 10; Gaps 5;
QY 557 AGTCAGACTATCATGATTACTGTGTTGACTGTGCTGCTATTGTGTAACCATGCC---- 613
DB 458 AGACAGACTATCTGATTGCTGCTTGCCTCTGCTGCTCATACAGAGCTACACCCACC 399
QY 613 --GATGTTAGTGAATTAGCGTCACTTGCCCTCAACCACTTGTGATACCAATTATCCT 670
DB 398 TTGCTTTGATAGTGTGAGTGTGCTGTACTTGCCACTGCCACTCAGGATTCATATTTC 339
QY 671 CACTGAATTGACATTTCTGCTTCACTGTCAAGCAATTCACAC-ATAATTCTGACCTACA 729
DB 338 CATGTATTTAATTTTGTAGTTGAGTGAAGTCACTGTGTTCCACTGTTAGATCTGACATAAC 279
QY 730 GAGGTGAGGATCATATAGCTCTTCAAGGATGCTGTGTAATCCCTCACAATTCATTCT 789

Db 278 AAGACGACCAATTACAGGGCTCTTCAAGATGCAGGTGCTGCCCTCACAAATCTGTTTCG 219
OY 790 CCTG-TTGTAGTGAAGAGTGCGCCCTCTGAGCCTCCAGGGTGGGTGTCAGGTC-ACA 847
Db 218 AAGACATGTGTCACAGGATATCTCTGATCCTCCCACTGTGTGAGTAGGCTTAA 159
OY 848 ATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTT-AAATCCCTCATGCTCAGT 906
Db 158 GTGACTATTCACACTCCACCATCCCAATCTCCCTAAGTCTTGGATTCCTTCCTACATT 99
OY 907 ACACGAGGCGAGGTCTAGCATTTTC 930
Db 98 AAACCAAGGAGACCAAGCATTTTC 75

RESULT 7
AI452907/c 475 bp mRNA EST 13-APR-1999
LOCUS
DEFINITION tJ46904.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2144598 3' similar to TR:002710 002710 GAG POLYPEPTIDE. ;
mRNA sequence.

ACCESSION AI452907
NID 94288636
VERSION AI452907.1 GI:4288636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE 1 (bases 1 to 475) Eutheria; Primates; Catarrhini; Homidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949381.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 693 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 455.

FEATURES
source
1.475
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19"
/clone="IMAGE:2144598"
/clone_1lb="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneIDs: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 143 a 85 c 125 g 122 t
ORIGIN

Query Match 8.9%; Score 99.6; DB 46; Length 475;

Best Local Similarity 62.5%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 134; Indels 10; Gaps 5;
OY 557 AGTCACTATCATGATGATGCTGTGAGTGTGCTGCTATGTGACTACCATGCC--- 613
Db 458 AGACAGACTATTCTGATGCTGCTTTGCTGCTGCTGCTATACAGAGCTACACCACC 399
OY 613 --GATGTTAGTGAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
Db 398 TTGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 339
OY 671 CACTGATGAGATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
Db 338 CATGTATTAAATTGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 279
OY 730 GAGTGAAGCATCATATAGCTCTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 789
Db 278 AAGACGACCAATTACAGGGCTCTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 219
OY 790 CCTG-TTGTAGTGAAGAGTGCGCCCTCTGAGCCTCCAGGGTGGGTGTCAGGTC-ACA 847
Db 218 AAGACATGTGTCACAGGATATCTTCTGATCCTCCCACTGTGTGAGTAGGCTTAA 159
OY 848 ATGATGAATGTATGATCGTGTCCCATTAACCAAGCCTTT-AAATCCCTCATGCTCAGT 906
Db 158 GTGACTATTCACACTCCACCATCCCAATCTCCCTAAGTCTTGGATTCCTTCCTACATT 99
OY 907 ACACGAGGCGAGGTCTAGCATTTTC 930
Db 98 AAACCAAGGAGACCAAGCATTTTC 75

RESULT 8
AA295609 245 bp mRNA EST 18-APR-1997
LOCUS
DEFINITION EST100780 Pancreas tumor 1 Homo sapiens cDNA 5' end similar to
Kallikrein, pancreatic, mRNA sequence.

ACCESSION AA295609
NID 91947944
VERSION AA295609.1 GI:1947944
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 245) Eutheria; Primates; Catarrhini; Homidae; Homo.
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.

TITLE
JOURNAL
MEDLINE
COMMENT Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On Nov 29, 1993 this sequence version replaced gi:635027.

Other_ESTS: THCL67057
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 384.

FEATURES

source

1. 429

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1085470"

/clone_1lb="NCI_CGAP_Gasl"

/tissue_type="gastric tumor"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: stomach; Vector: Bluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pooled gastric tumors. 5' adaptor sequence: 5'

GAAATCGGCACGAG 3' 3' adaptor sequence: 5'

CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.0 kb."

BASE COUNT 114 a 100 c 78 g 137 t

ORIGIN

Query Match 7.3%; Score 82.2; DB 35; Length 429;

Best Local Similarity 62.0%; Pred. No. 8.3e-14;

Matches 199; Conservative 0; Mismatches 113; Indels 9; Gaps 4;

QY 532 ACTAGCCAGCACCAGTAGTCT-CCGAGTGCAGCTATCATGATGCTGTGACTGTGC 590

DB 109 ATTAGCCAAATGCCAGAGCTCTGCACAAATCAGACTGTTCTGATGCTCTTGGCAGTGT 168

QY 591 TGTCTATTGTACTAACCATGCGGATGTTT-----AGTGAAATTAAGCTGCTGAGTGGCT 644

DB 169 TGTCCATTACAGTAGTACACACCCCTTTCTTCTTGTGATGTTGAGTGCCTCCAGCTGGCCC 228

QY 645 CAACCATCTTGGTATCCAGTATCTCCTCAGTGAATGAGATTTCCTGCTTCAAGATGCT 704

DB 229 CTGCCACCTCAGAGATTCATATTTCCATGTATTTAATTTTGTAGTGAAGTGAAGTGA 288

QY 705 ATCCACAC-ATAATTTCTGACCTACAGAGTGAAGGATCATATAGCTCTTCAAGATGCT 763

DB 289 GTTCCCACTGTAGACCTGGCATACAGAGAAAGCAATTACAGGCTCTTCAAGATGCA 348

QY 764 GGTACTCCCTCACAATTCATTTCTCTGTGTAG-TGAAGGTGCGCCCTCTGAGGCC 822

DB 349 GATGATGCCCTTACAAATCTATTTCACAAAGATATGATCAAAAGTTATGCTCTTGAGCCC 408

QY 823 TCCAGAGGTGGGTGTCAGGT 843

DB 409 TCCAGAGGTGGGTGTAAGT 429

RESULT 11

AA776018 407 bp mRNA EST 05-FEB-1998

LOCUS ae83g11.s1 Strata gene schizo brain S11 Homo sapiens cDNA clone

DEFINITION IMAGE:970820 3', mRNA sequence.

ACCESSION AA776018

NID 92835352

VERSION AA776018.1 GI:2835352

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS 1 (bases 1 to 407)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacey, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

On Jan 17, 1998 this sequence version replaced gi:2044645.

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available royalty-free through LBNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: polyT not found

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

1. 407

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:970820"

/clone_1lb="Stratagene schizo brain S11"

/sex="male"

/tissue_type="schizophrenic brain S-11 frontal lobe"

/dev_stage="34 years old"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Library

constructed from S-11 frontal lobe, male, 34 years old,

50% caucasian, 50% Aleutian. Schizophrenic suicide.

Random primed into EcoRI site of ZAP II Vector. Mass

excised. Custom library. Avg insert length 1.4kb.

Material obtained by Johnston N., Torrey, E.F., Yolken R.,

and the Stanley Neuropathology Consortium - Analysis of

RNAs from the Brains of Individuals with Psychiatric

Diseases (Unpublished) Stanley Neurovirology Laboratory,

Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 120 a 73 c 98 g 116 t

ORIGIN

Query Match 7.3%; Score 81.4; DB 38; Length 407;

Best Local Similarity 57.3%; Pred. No. 1.4e-13;

Matches 207; Conservative 0; Mismatches 146; Indels 8; Gaps 3;

QY 495 ACTACTGCTCTGATCATCACCAGGACACATGATCACTAGCCAGCACCATAGT-TCTC 553

DB 382 ATTCCTATCAATTTCACTCTAGGACACCATGATTAATTAACCAATGCCGTAGTCTGAC 323

QY 554 CGAAGTCAGACTATCATGATTTACTGTGTGACTGTGCTGTCTATTGTACTAACCATGCCG 613

DB 322 ACGAGTCAGACTATTCGATTTGCTGCTTGCCTGTGCTGCTTACAGTAGCTACACC 263

QY 614 ATGTTAGTGAAT-----TAGCGTCACTTGCCCTCAACCATCTGTATCCAGTAT 667

DB 262 ACCTTACCTTGAAGGTGACTGCCACCAATTGGCCCCCTGCCACCTCTGATTAATTTAT 203

QY 668 CCTCACTGAATGAGATTTCCTGCTTCAAGTGTACAGCCATTCACATATTTCTGACCTA 727

DB 202 TCCCATTTGATTTAATTTTATAGTTAGTGAC-TAGTTCACAGTGTAGATCTGACATA 144

QY 728 CAGAGGTAGGATCATATAGCTCTTCAAGGATGCTGTACTCCCTCACAATTCATTT 787

DB 143 AGGAAAGAGCAATTACAGGCTCTCTCAAAAATGACAGGTGCTGCTCTCACAATTCATTT 84

QY 788 CTCCTGTTGAGTGAAGGTGCGCCCTCTGAGCCCTCCAGGGTGGGTGTGACAGTACA 847

DB 83 CGCAAGCATTTGTTAAGAGTGTATCTTACAGCCCTCCAGCTGGATGAGTAGTCTTAA 24

QY 848 A 848

DB 23 A 23

RESULT 12

AA542994 523 bp mRNA EST 19-AUG-1997

LOCUS n155b01.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:980713

DEFINITION similar to TR:G1020091 G1020091 NEUROPSIN. ;, mRNA sequence.

ACCESSION AA542994

NID 92291474

VERSION AA542994.1 GI:2291474

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405040.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1057 Std Error: 0.00
Seq primer: -21m13 forward (Amersham)
High quality sequence stop: 462.

FEATURES
Location/Qualifiers

1.523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:980713"
/clone_lib="NCI-CGAP_Ov2"
/sex="female"
/tissue_type="ovary"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from invasive ovarian
tumor, CDNA made by oligo-dT priming. Non-directionally
cloned. Size-selected on agarose gel, average insert size
600 bp. Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT 112 a 171 c 138 g 102 t
ORIGIN

Query Match 7.2%; Score 81; DB 35; Length 523;
Best Local Similarity 53.7%; Pred. No. 2e-13;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

OY 6 CTCGAGCCCTGGCAGGCGCACTGGTCATGAAAAAGATTGTTCTGCTGGGCGTCT 65
DB 121 CTCACACCATGGCAGGTGGCCCTGCTCAGTGCATCAGCTCCACTGCGAGGCGTCT 180
OY 66 GGTGATCCGCACTGGGTGCTGACGCCGACACTGTTCAGAACTCCTACACCATCGG 125
DB 181 GGTCAATGAGCGCTGGGTGCTCTACTGCCGCCCACTGCAGATGAATGATACACCTGCA 240
OY 126 GCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAG 185
DB 241 CCTGGGCACTGATACGCTGGGCGACAGGAGAG-----CTCAGAGGATCAAGGCCCTC 291
OY 186 CCTCTCCGTACGGCACCAGAGTACAAGACCCCTGCTGCTAACGACCTCATGCTCAT 245
DB 292 GAAGTCATTCGCCACCCCGGCTACTCCACACAGACCCCATGTATGACCTCATGCTCGT 351
OY 246 CAAGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGAGAGCATCAGCATGCTTCGCA 305
DB 352 GAAGCTCAATAGCCAGGCGAGGCTGTATCCATGTTGAAGAAAGTCAAGGCTGCCCTCCG 411
OY 306 GTGCCCTACCGCGGGGAACCTCTGCTCTGCTGGCTGGGGTCTGCTGGCGACGATGC 365
DB 412 CTGGGACCCCTGGAACCACTGTACTGTCTCCGGCTGGGGCACTACCAAGAGCCCA 471

OY 366 TGTGA 370
DB 472 TGTGA 476

RESULT 13
AA535837 486 bp mRNA EST 21-AUG-1997
LOCUS n179g05.s1 NCI-CGAP_Pr10 Homo sapiens CDNA clone IMAGE:998744
DEFINITION similar to gb:S39329 GLANDULAR KALLIKREIN 2 PRECURSOR (HUMAN);,
mRNA sequence.

ACCESSION AA535837
NID 92280090
VERSION AA535837.1 GI:2280090
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 486)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Apr 14, 1993 this sequence version replaced gi:315330.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1091 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 385.

FEATURES
Location/Qualifiers

1.486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:998744"
/clone_lib="NCI-CGAP_Pr10"
/sex="male"
/tissue_type="invasive prostate tumor"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMPI0; mRNA made from
invasive prostate tumor, CDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

BASE COUNT 96 a 142 c 145 g 103 t
ORIGIN

Query Match 6.8%; Score 76.2; DB 35; Length 486;
Best Local Similarity 66.8%; Pred. No. 5.1e-12;
Matches 125; Conservative 0; Mismatches 58; Indels 4; Gaps 1;

OY 340 GGCTGGGCTCTGCTGGGGAAGATGCTGTGATGCCATCCAGTCCAGACTGTGGAGGC 399
DB 58 GCCTGTCTGTGGGTGCACTGTGCGCTGCCCTCATCCAGTCTCGAATGTGGAGGC 117
OY 400 TGGAGTGTGAGAAGCTTCCCAACCTGGCAGG---GTGTACCATTTCCGCACTTC 455
DB 118 TGGAGTGTGAGAAGCATTTCCCAACCTGGCAGGTGGCTGTGTACAGTATGATGGCA 177
OY 456 CAGTGAAGAGACGTCTGTGATCTCTACTGGGTGCTCAGTGTGATGCTGATGATGATGAT 515

| | | | | | | |
|----|-----|-----|-----------------------------|-----------------------------|-------|-----|
| Db | 178 | CAC | TGTGGGGGGTGTCTCTGGTGCACCCCA | GTGGGTGCTCACAGCTGCCCATTTGCC | TAAAG | 237 |
| QY | 516 | CGG | AACA | 522 | | |
| | | | | | | |
| Db | 238 | AAG | AATA | 244 | | |

| RESULT | 14 |
|------------|--|
| AA579026 | |
| LOCUS | 572 bp mRNA |
| DEFINITION | EST 12-SEP-1997 |
| | nf34f02.s1 NCI-CGAP_Prl Homo sapiens CDNA clone IMAGE:915675 |
| | similar to gb:S39329 GLANDULAR KALLIKREIN 2 PRECURSOR |
| | (HUMAN); contains Alu repetitive element; mRNA sequence. |

| | | |
|-----------|--------------|------------|
| ACCESSION | AA579026 | |
| NID | 92357210 | |
| VERSION | AA579026.1 | GI:2357210 |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 572) |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | On Sep 12, 1996 this sequence version replaced gi:1407214. |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbip/image/image.html

Insert Length: 921 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 159.

Location/Qualifiers
1. 572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915675"
/clone_1lb="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)/17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into PAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman"

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 118 a | 153 c | 181 g | 120 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 6.8%; | Score 76.2; | DB 35; | Length 572; |
| Best Local Similarity | 66.8%; | Pred. No. 5.5e-12; | | |
| Matches 125; Conservative | 0; | Mismatches 58; | Indels 4; | Gaps 1; |

[illegible]

| | | | |
|----|-----|--|-----|
| Db | 16 | GCCCTGCTCTGTGGGGTGCACTGGTGGCCGTGCCCCCTCATCCAGTCTCGGATTGTGGAGGC | 75 |
| OY | 400 | TGGAGTGTGAGAAAGCTTTCACCAACCTGGCAGG---GTGTACCATTTCGGCACTTC | 455 |
| | | | |
| Db | 76 | TGGAGTGTGAGAAAGCATTTCCCAACCTGGCAGGTGGCTGTGTACAGTCATGGATGGCA | 135 |
| OY | 456 | CAGTGCAGGACGTCCTGCTGCATCTCACTGGGTGCTCACTGCTCACTGCATGCATCACC | 515 |
| | | | |
| Db | 136 | CACGTGTGGGGGTCTCTGTGTCAACCCCAGTGGGTCTACAGCTGCCCATTTGCTTAAG | 195 |
| OY | 516 | CGGACA | 522 |
| | | | |
| Db | 196 | AAGATA | 202 |

| | | | |
|------------|--|-----------|------|
| RESULT | 15 | | |
| T29510 | | | |
| LOCUS | T29510 | 253 bp | mRNA |
| DEFINITION | EST82448 Human Prostate gland Homo sapiens CDNA 5' end similar to kallikrein 1, renal/pancreas/salivary (HT:711), mRNA sequence. | | |
| ACCESSION | T29510 | | |
| NID | 9611608 | | |
| VERSION | T29510.1 | GI:611608 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |

| REFERENCE | 1 (bases 1 to 253) |
|-----------|---|
| AUTHORS | Adams, M.D., Kerlavsky, A.R., Fleischmann, R.D., Fuldner, R.A., |

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M., Klimek, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Li, H., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence

Nature 377, 3-174 (1995)

| | |
|---------|---|
| TITLE | Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence |
| JOURNAL | Nature 377, 3-174 (1995) |
| MEDLINE | 96026280 |

COMMENT

Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| SOURCE | 1. .253 |

```

/organism="Homo sapiens"
/db_xref="ATCC (inhost):106507"
/db_xref="taxon:9606"
/clone_lib="Human Prostate gland"
/note="Organ: prostate gland"

```

BASE COUNT 42 a 68 c 81 g 59 t 3 others

ORIGIN

Query Match 6.7%; Score 74.8; DB 20; Length 253;

Best Local Similarity 65.8%; Pred. No. 1e-11;
Matches 123; Conservative 0; Mismatches 60; Indels 4; Gaps 1;

```
OY 340 GGCCTGGGGTCTGCTGGCGAAGCATGCTGTGATGCCATCCAGTCCAGACTGTGGAGGC 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 GGCCTGTCTGTGGGGTGCACACTGTGCGCGTCCCTCATCCAGTCTCGGATTGTGGAGGC 98

OY 400 TGGAGTGTGAGAAGCTTCCCAACCCCTGGCAGG---GTGTACCATTTGGCACTTC 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 TGGAGTGTGAGAAGCATTCACACCCCTGGCAGGTGCTGTGTACAGTCATGATGGCA 158

OY 456 CAGTCAAGAGACGCTCCTGCTGCATCCTCACTGGGTGCTCACTACTGCTCACTGCATCACC 515
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 CACTGTGGGGNTNTCCGTGTGCAACCCCAAGTGGGTGCTCAACAGCTGCCCATGTGNTAAAG 218

OY 516 CGGAACA 522
    |||||
Db 219 AAGAATA 225
```

Search completed: September 25, 1999, 12:05:32
Job time: 8136 sec

OM of: US-09-030-606-223 to: A_Geneseq_36:* out_format : pfs
Date: Sep 25, 1999 11:36 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL-frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29804/app-query.fasta.1
-DB-A_Geneseq_36-QFMT-fastan -SUFFIX-rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-biosum62
-TRANS-human40.cdl -LIST=45 -DOCCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-223
Query length: 383
Database: A_Geneseq_36:.*
Database sequences: 188963
Search time (sec): 185.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|---------------------|------|-------|--------|--------|------|----------------------------------|
| A_Geneseq_36:Y07044 | - | 64.50 | 147.96 | 0.3541 | 258 | Breast cancer associated antige |
| A_Geneseq_36:W34260 | + | 61.50 | 127.42 | 1.43 | 892 | Rat ob receptor isoform c'. New |
| A_Geneseq_36:W37338 | + | 61.50 | 127.40 | 1.43 | 894 | Ob protein receptor. ob protein |
| A_Geneseq_36:W37337 | + | 61.50 | 127.40 | 1.43 | 894 | Ob protein receptor. ob protein |
| A_Geneseq_36:W34258 | + | 61.50 | 127.39 | 1.43 | 895 | Rat ob receptor isoform f. New |
| A_Geneseq_36:W34259 | + | 61.50 | 126.13 | 1.48 | 1015 | Rat ob receptor isoform g. New |
| A_Geneseq_36:W23399 | + | 61.50 | 124.77 | 1.54 | 1162 | Rat ob receptor (fa mutant). R |
| A_Geneseq_36:W23398 | + | 61.50 | 124.77 | 1.54 | 1162 | Rat ob receptor (wild-type). R |
| A_Geneseq_36:W34257 | + | 61.50 | 124.77 | 1.54 | 1162 | Rat wild-type ob receptor. New |
| A_Geneseq_36:W75958 | + | 57.50 | 131.64 | 3.71 | 200 | Mouse cell surface protein. Cel |
| A_Geneseq_36:R58650 | + | 58.00 | 126.47 | 3.76 | 383 | Fowlpox virus (FPV) genome Ecd |
| A_Geneseq_36:R58654 | + | 58.00 | 126.47 | 3.76 | 383 | Fowlpox virus protein encoded b |
| A_Geneseq_36:W10692 | + | 58.00 | 126.47 | 3.76 | 383 | Homology vector 443-88.8-encode |
| A_Geneseq_36:W01213 | - | 56.50 | 134.15 | 4.52 | 119 | Serine protease PfSP19-119. DNA |
| A_Geneseq_36:P92108 | + | 57.50 | 120.01 | 5.17 | 637 | Human growth hormone receptor. |
| A_Geneseq_36:P81326 | + | 57.50 | 120.00 | 5.17 | 638 | Human growth hormone receptor. |
| A_Geneseq_36:W33394 | + | 57.50 | 120.00 | 5.17 | 638 | Human growth hormone receptor. |
| A_Geneseq_36:W41988 | - | 56.50 | 126.58 | 5.61 | 253 | Flea serine protease SEQ ID NO: |
| A_Geneseq_36:W98665 | + | 56.50 | 125.78 | 5.73 | 274 | H. pylori GHPD 489 protein. New |
| A_Geneseq_36:W74900 | + | 54.50 | 135.46 | 7.45 | 61 | Human secreted protein encoded b |
| A_Geneseq_36:W20224 | + | 55.50 | 124.93 | 7.69 | 228 | H. pylori transporter protein, |
| A_Geneseq_36:W20668 | + | 55.50 | 124.37 | 7.81 | 241 | H. pylori transporter protein c |
| A_Geneseq_36:W01144 | + | 55.00 | 128.76 | 7.88 | 136 | Mab 1.4 heavy chain, directed a |
| A_Geneseq_36:W44169 | + | 55.00 | 128.76 | 7.88 | 136 | Monoclonal antibody 1.4 heavy c |
| A_Geneseq_36:R77604 | + | 56.50 | 107.61 | 9.64 | 1676 | Pro-C5 polypeptide. Treating g |
| A_Geneseq_36:W27838 | - | 55.00 | 120.69 | 9.87 | 298 | Staphylococcus aureus protein c |
| A_Geneseq_36:Y07979 | - | 54.50 | 125.54 | 9.89 | 164 | Human secreted protein fragment |
| A_Geneseq_36:W01808 | + | 55.00 | 118.78 | 10.49 | 1065 | Arabidopsis cellulose biosynth |
| A_Geneseq_36:R89906 | + | 54.50 | 116.05 | 12.96 | 422 | Human papillomavirus 6b E6/E5a/ |
| A_Geneseq_36:P70567 | + | 53.00 | 127.07 | 14.16 | 94 | Product of ORF 9 from MYMV compl |
| A_Geneseq_36:W20517 | + | 53.50 | 116.48 | 16.76 | 309 | H. pylori surface or membrane p |
| A_Geneseq_36:W21030 | + | 53.50 | 116.45 | 16.77 | 310 | H. pylori transmembrane protein |
| A_Geneseq_36:W36811 | - | 54.00 | 111.27 | 17.00 | 594 | Human GDP-fucose pyrophosphoryl |
| A_Geneseq_36:W00551 | - | 53.50 | 115.29 | 17.34 | 348 | Human histamine H2 receptor pol |
| A_Geneseq_36:W31712 | - | 53.50 | 115.29 | 17.34 | 348 | Protein sequence of a region of |
| A_Geneseq_36:Y00889 | - | 53.50 | 114.98 | 17.49 | 359 | Human histamine H2 receptor pro |
| A_Geneseq_36:W93949 | + | 53.50 | 113.14 | 18.43 | 431 | Human regulatory molecule HRM-5 |
| A_Geneseq_36:W26733 | + | 54.00 | 108.28 | 18.52 | 801 | Staphylococcus carnosus nitrite |
| A_Geneseq_36:W68398 | + | 53.50 | 112.69 | 18.68 | 451 | Clostridium botulinum type D to |
| A_Geneseq_36:W19105 | + | 53.00 | 115.10 | 19.94 | 310 | Rat pheromone receptor VN4. Nuc |
| A_Geneseq_36:W98010 | - | 53.50 | 109.40 | 20.52 | 626 | BCG Ag85b antigen-phagolysomal |

A_Geneseq_36:W89953 + 52.00 119.82 22.80 148 ! Antigen from cluster 54. New
A_Geneseq_36:W19106 + 52.50 114.85 22.98 278 ! Rat pheromone receptor VN5.
A_Geneseq_36:R31610 + 51.00 124.35 26.20 72 ! Fragment XXVI with homology t

seq_name: A_Geneseq_36:Y07044

seq_documentation_block:
ID Y07044 standard; Protein; 258 AA.
AC Y07044;
DT 02-JUL-1999 (first entry)
DE Breast cancer associated antigen precursor sequence.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PE 15-JUL-1998; U14679.
PF 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI fireundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Disclosure: Page 428-429; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 258 AA:

alignment_scores:
Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 1
Percent Similarity: 50.000 Percent Identity: 23.684

alignment_block:

US-09-030-606-223/rev x Y07044 ..

Align seg 1/1 to: Y07044 from: 1 to: 258

292 CCCCCAACAAAGTTAAATATATATGTCATTCACAGGAATCAAAATCTTT 243
||| ::|||::|||:: |||::: |||::: |||:::
14 ProHISGLYArgLeuGIuTYrThcGIuHISGLNGlyArgIleLysASnaI 30
242 TAGAATAGCACACTCCAAACAAGTGAKGGACACACTAATTCCTTAGA 193
||| |||||::: |||||::: |||||::: |||||:::
30 aArgGLuAlaHISserGln..... 36
192 CTTCCTTGGCAGCATTTACTTTGATAAGAAGTCTCCAAATAAATACAA 143
37:|||||::: |||||::: |||||::: |||||:::
142 AATTTGGCAGACATTTTAATCTTGTCAGACACAATGTAAAGAAATGCC 93


```

:::
747 LeuSerSer...SerCysValIleLeu.SerTyrThrLeuSerProAsnA 762
:::|||||
266 ACAATTATATTTTAACTTTGGTGGCGGA 294
|| ::||| ||||| |||
762 sPTyrSerLeuLeuTyrLeuValIleGlu 771
```

seq_name: A_Geneseq_36:W37337

seq_documentation_block:

ID W37337 standard; Protein; 894 AA.
AC W37337;
DT 07-MAY-1998 (first entry)
DE Ob protein receptor.
KW Ob protein receptor; obese phenotype detection; rat.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc_difference 805 /note="encoded by TGT"
PN WO9741217-A1.
PD 06-NOV-1997.
PE 24-APR-1997; J01470.
PR 30-APR-1996; JP-134422.
PA (SAKA) OTSUKA PHARM CO LTD.
PI Iida M, Kodaira T, Murakami T, Shima K;
DR WPI; 97-549722/50.
DR N-PSDB; T97149.
PT Ob protein receptor gene and variant having cytosine at base 806 -
PT for diagnosis of mutation(s) leading to development of obesity in
PT animals
PS Claim 1; Page 49-54; 71pp; Japanese.
CC This sequence represents the obesity protein receptor of the
CC invention. The gene sequence may be used for screening of warm blooded
CC animals (such as rats) for spontaneous mutations resulting in the variant
CC Ob protein receptor gene which leads to animals having an obese
CC phenotype. These are useful in the study of mechanisms of obesity and the
CC development of anti-obesity drugs.
SQ Sequence 894 AA;

alignment_scores: Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:

US-09-030-606-223 x W37337 ..

Align seg 1/1 to: W37337 from: 1 to: 894

```

17 AAAAACAATTTCT.....TCATTCAGAAAAATATCTTA..... 49
|||||::||| |||::| ::::|
667 LysAsnAspSerLeuCysSerValArgArgTyrValValLysHisArgTh 683
50 .....GGGACT.....GATATGGTAATTATGTCATTTTA 80
||||| ||||| |||::| ||||| ::::| |||||
683 GAlaHisAsnGlyThrTyrSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTRTTKTGGGGCAATTTCCTTACAT.....TGCTTGA 115
|||||::||| ||| ||| ::::|
700 ThrPheLeuTyrPalaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGTCGTGTCGCAAAATTTGTATTTTATTTGAGACTTCT 165
:::|||| ||||| ::||| ::||| ::||| ::|||
716 nSerIleGlyAlaSerLeuValAsnPhAsnLeuThrPheSerTyrProM 733
166 TATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAGTGTCTCC 215
:::||||| ||| ::||| ::||| ::||| ::|||
733 eSerLysVal.....AsnAlaValGlnSerLeuSerAlaTyrPro 746
216 MTGACTGTGTTGGAGTGTGCTATCTTAAAGATTGATTGCTCTGAATG 265
::: |||||::||| ||||| ::||| ::||| |||||
747 LeuSerSer...SerCysValIleLeu.SerTyrThrLeuSerProAsnA 762
```

```

266 ACAATTATATTTTAACTTTGGTGGCGGA 294
|| ::||| ||||| |||
762 sPTyrSerLeuLeuTyrLeuValIleGlu 771
```

seq_name: A_Geneseq_36:W34258

seq_documentation_block:

ID W34258 standard; Protein; 895 AA.
AC W34258;
DT 11-MAY-1998 (first entry)
DE Rat ob receptor isoform f.
KW Ob receptor isoform f; obesity; leptin; ligand; rat.
OS Rattus sp.
PN WO9742340-A1.
PD 13-NOV-1997.
PE 02-MAY-1997; U07521.
PR 24-MAY-1996; GB-010995.
PR 06-MAY-1996; US-016899.
PA (MERI) MERCK & CO INC.
PI Caskey CT, Hess JW, Hey P, Phillips MS;
DR WPI; 97-558993/51.
DR N-PSDB; T93104.
PT New isoform(s) of the Ob (leptin) receptor - used for identifying
PT specific binding ligands, potentially useful for study, prevention
PT and treatment of obesity
PS Disclosure; Page -: 34pp; English.
CC This protein comprises isoform f of the rat ob receptor. The ob
CC receptor has numerous isoforms resulting from alternative splicing;
CC 3 novel isoforms, designated f, g and c' are disclosed (see
CC W34258-60). Isoform f differs from the 1162-residue wild-type
CC protein (see W34257) in that after Lys-889 there are only 6 amino
CC acids. In the cDNA (see T93104), these codons are then followed by
CC a stop codon. The f isoform can be expressed in host cells,
CC particularly E. coli, yeast or mammalian cells. It is used to
CC identify specific binding ligands. Agonists, antagonists and
CC ligand mimetics can be identified that are potentially useful in
CC the study, prevention and treatment of obesity.
CC (NB. the amino acid sequence of isoform f was produced by
CC adaptation of the wild-type ob receptor sequence provided in Fig 1
CC of the specification).
SQ Sequence 895 AA;

alignment_scores: Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:

US-09-030-606-223 x W34258 ..

Align seg 1/1 to: W34258 from: 1 to: 895

```

17 AAAAACAATTTCT.....TCATTCAGAAAAATATCTTA..... 49
|||||::||| ||| ||| ::::|
667 LysAsnAspSerLeuCysSerValArgArgTyrValValLysHisArgTh 683
50 .....GGGACT.....GATATGGTAATTATGTCATTTTA 80
||||| ||||| |||::| ||||| ::::| |||||
683 GAlaHisAsnGlyThrTyrSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTRTTKTGGGGCAATTTCCTTACAT.....TGCTTGA 115
|||||::||| ||| ||| ::::|
700 ThrPheLeuTyrPalaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGTCGTGTCGCAAAATTTGTATTTTATTTGAGACTTCT 165
:::|||| ||||| ::||| ::||| ::||| ::|||
716 nSerIleGlyAlaSerLeuValAsnPhAsnLeuThrPheSerTyrProM 733
166 TATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAGTGTCTCC 215
:::||||| ||| ::||| ::||| ::||| ::|||
733 eSerLysVal.....AsnAlaValGlnSerLeuSerAlaTyrPro 746
```



```

216 MTRCAGITGTTGGAGTGTGCATTCCTAAAGAATTGGATTCCCTCGAATG   265
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
747 LeuSerSer...SercysValIleLeu..SerTriPThrLeuSerProAsnA   762
266 ACAATTATATTTTAACCTTTGGTGCGGGA      294
    ||::::||| |||||| |||
762 spTyrrSerLeuLeuTyrrLeuValIleGlu     771
```

seq_name: A_Geneseq_36:W34259

seq_documentation_block:
ID W34259 standard; Protein; 1015 AA.
AC W34259;
DT 11-MAY-1998 (first entry)
DE Rat ob receptor isoform g.
KW Ob receptor isoform g; obesity; leptin; ligand; rat.
OS Rattus sp.
PN M09742340-A1.
PD 13-NOV-1997.
PF 02-MAY-1997; U07521.
PR 24-MAY-1996; GB-010995.
PR 06-MAY-1996; US-016899.
PA (MERI) MERCK & CO INC.
PI Caskey CT, Hess JW, Hey P, Phillips MS;
DR MPI; 97-558993/51.
PT New isoform(s) of the Ob (leptin) receptor - used for identifying
PT specific binding ligands, potentially useful for study, prevention
PT and treatment of obesity
PS Disclosure; Page -: 34pp; English.
CC This protein comprises isoform g of the rat ob receptor. The ob
CC receptor has numerous isoforms resulting from alternative splicing;
CC 3 novel isoforms, designated f, g and c' are disclosed (see
CC W34258-60). Isoform g is much shorter than the 1162-residue
CC wild-type sequence (see W34257). 18 N-terminal amino acids are
CC spliced to a large fragment of the wild-type molecule, beginning at
CC Pro-166. The isoform then extends for the remainder of wild-type
CC ob receptor molecule. Alternatively, it could contain another
CC isoform variation, such as isoform a, b, c, d, e or f. The g
CC isoform can be expressed in host cells, particularly E. coli, yeast
CC or mammalian cells. It is used to identify specific binding
CC ligands. Agonists, antagonists and ligand mimetics can be
CC identified that are potentially useful in the study, prevention and
CC treatment of obesity.
CC (NB, the amino acid sequence of isoform g was produced by
CC adaptation of the wild-type ob receptor sequence provided in Fig 1
CC of the specification).
SQ Sequence 1015 AA;

```
alignment_scores:
  Quality: 61.50
  Ratio: 0.992
  Percent Similarity: 55.856
  Length: 111
  Gaps: 8
  Percent Identity: 31.532
```

alignment_block:

US-09-030-606-223 x W34259

Align seg 1/1 to: W34259 from: 1 to: 1015

17 AAAAACAAATTC...TCATTCAGAAAAATTATCTTA..... 49
 |||||:::|||| || ||::: :::::
 520 LysAsnAspSerLeuCysSerValArgArgTyrValValLysHisArgTh 536
 50GGGACT.....GATATTGCTAATTATGTCATTTA 80
 ||||| ||:::||||| :::::|||||
 536 AlaHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 552
 81 ATWRTTRTKTGGGGCATTCCCTTACAT.....TGTCTTGA 115
 ||::: ||| ||| :::::
 553 ThrPheLeuTrpAlaGluSerAlaHisThrValThrValLeuAlaIleAs 569
 116 CAAAGATTAAATGCTCTGTGCCAAAATTTTGATTTTATTTGGAGACTTCT 165

```

:::|||| |::: ::||| ::: ||::: ;
nSerIleGlyAlaSerLeuValasnPheasnLeuThrPheserTrpRom 586

166 TATCAAAAGTATATGCTGCCAAGAAGCTTAAGGAATTAGTAGTGTTCCC 215
   :||||||||| ::: ||| :::::||::: |||
586 etserLySval.....AsnaLaVaGlInSerLeuSerLatyrPro 599

216 MTCAC TTGTGGAGTGTGCtATrCTAAAAGATT TGATTTCCtCGAAtG 265
   ::: :::||||:||||| ::: :::|||| |!!!
600 LeuSerSer..SerCysValIleLeu.SerrTpThrLeuSerProAsna 615

266 ACAATTATATTTTACTTTGGTGGGGGAA 294
   || :::|||| |!!!! |!!
615 sPTyrsErLeuleuTyrrLeuValIleglu 624
```

seq_name: A_Geneseq_36:W23399

```

seq_documentation_block:
ID. W23399 standard; Protein; 1162 AA.
AC W23399;
DT 23-FEB-1998 (first entry)
DE Rat ob receptor (fa mutant).
KW Ob receptor; obesity; leptin; rat; rodent; animal model; ligand;
KM fatty; fa mutation; therapy.
OS Rattus sp.
FH Key
FT Peptide
FT 1..28
FT /label= Sig_peptide
FT 29..1162
FT /label= Mat_protein
FT 29..830
FT /label= Extracellular_domain
FT 839..860
FT /label= Transmembrane_domain
FT /note= "6"
FT 860..1162
FT /label= Cytoplasmic_domain
FT 269
FT Misc_difference
FT /note= "residue 269 is Gln in wild-type OB-R"

```

PD 28-AUG-1997.
PF 18-FEB-1997; U02397.
PR 25-APR-1996; GB-008473.
PR 22-FEB-1996; US-090405.
PR 22-MAR-1996; US-013969.
PA (MERI) MERCK & CO INC.
PI Caskey CT, Hess JW, Liu Q, Phillips MS;
DR WPI; 97-435085/40.
DR N-PSDB; T64962.
PT Rat wild-type and mutant ob receptor protein - useful in
PT identification of new ligands for prevention and treatment of
PT obesity
PT Claim 4; page -; 35pp; English.
CC This protein comprises the rat fatty (fa) mutant ob receptor
CC (OB-R). Its amino acid sequence was deduced from a cDNA sequence
CC (see T64962) obtained by PCR amplifications. The sequence differs
CC at only 1 position from the wild-type (i.e. lean) OB-R sequence;
CC the fa OB-R mutant receptor has proline at amino acid 269 rather
CC than glutamine (see W23399). This is a result of an A to C
CC transversion in the fa OB-R gene (see T64962). Every tissue
CC examined in an fa/fa Zucker rat was homozygous for this mutation.
CC The wild-type and mutant OB-R can be expressed in prokaryotic or
CC eukaryotic host cells. They are useful in assays to identify
CC putative ligands, which may be receptor agonists or antagonists or
CC bind with little effect, and which may be useful in understanding
CC obesity and for its prevention and treatment.
CC (NB. the sequence for the fa OB-R was obtained by adapting the
CC sequence of wild-type OB-R provided in Fig 1 of the specification).
CC Sequence 1162 AA;
SQ

```
alignment_scores:
  Quality: 61.50
  Length: 111
```


Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:

US-09-030-606-223 x W23399 ..

Align seg 1/1 to: W23399 from: 1 to: 1162

```
17 AAAAACAATCT.....TCATTCAGAAAAATATCTTA..... 49
|||||:||||| ||| |||:| :|||:
667 LysAsnAspSerLeucysSerValArgArgTyrValVallySHisArgTh 683
50 .....GGGACT.....GATATGGTAATTAAGTCAATTGA 80
||||| |||:| |||:| |||:| |||:|
683 rAlaHisAsnGlyThrTyrSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTTCTKGGGGCATTTCTTACAT.....TGCTCTGA 115
|||:| ||| ||| |||:| :|||:|
700 ThrPheLeuTyrPalagIuSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGCTGTGCGCAAAATTTGTATTTTATTTGAGACTTCT 165
:|:|:| |||:| :|||:| :|||:|
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTyrProM 733
166 TATCAAAAGTAATGCTGCCAAGAAGCTCTAAGGAATAGTAGTCTCC 215
:|:|:| ||| ||| |||:| :|||:|
733 etSerIysVal.....AsnAlaValGlnSerLeuSerAlaTyrPro 746
216 MTCACCTGTTTGAGAGTGCTCTATCTAAAGATTGATTCCTGGAATG 265
:|:|:| |||:| :|||:| |||:| |||:|
747 LeuSerSer...SerCysValIleLeu.SerTyrThrLeuSerProAsnA 762
266 ACAATTATATTTTAACCTTTGGTGGGGGAA 294
|| :||| ||| ||| |||
762 sPTyrSerLeuLeuTyrLeuValIleGlu 771
```

seq_name: A_Geneseq_36:W23398

seq_documentation_block:

ID W23398 standard; Protein; 1162 AA.

AC W23398;

DT 23-FEB-1998 (first entry)

DE Rat ob receptor (wild-type).

KW Ob receptor; obesity; leptin; rat; rodent; animal model; ligand;

KW fatty; fa mutation; therapy.

OS Rattus sp.

Key Location/Qualifiers

FT Peptide 1..28

FT Protein /label= sig_peptide

FT Protein 29..1162

FT Domain /label= Mat_protein

FT Domain 29..830

FT Domain /label= Extracellular_domain

FT Domain 839..860

FT Domain /label= Transmembrane_domain

FT Domain /note= "6"

FT Domain 860..1162

FT Misc_difference 269

FT /label= Cytoplasmic_domain

FT /note= "site of mutation in fa OB-R"

PN WO9731015-A1.

PD 28-AUG-1997.

PF 18-FEB-1997; U02397.

PR 25-APR-1996; GB-008473.

PR 22-FEB-1996; US-090405.

PR 22-MAR-1996; US-013969.

PA (MERI) MERCK & CO INC.

PI Caskey CT, Hess JW, Liu Q, Phillips MS;

DR N-PSDB; T64961.

DR N-PSDB; T64961.

PT Rat wild-type and mutant ob receptor protein - useful in

PT Identification of new ligands for prevention and treatment of

PT obesity

PS Claim 5; Fig 1A-1B; 35pp; English.
CC This protein comprises rat wild-type (i.e. lean) ob receptor
CC (OB-R). Its amino acid sequence was deduced from a cDNA sequence
CC (see T64961) obtained by PCR amplifications. The sequence differs
CC at only 1 position from the OB-R sequence from rats with a fatty
CC (fa) mutation; the fa OB-R mutant receptor has proline at amino
CC acid 269 rather than glutamine (see W23399). This is a result of
CC an A to C transversion in the fa OB-R gene (see T64962). Every
CC tissue examined in an fa/fa Zucker rat was homozygous for this
CC mutation. The wild-type and mutant OB-R can be expressed in
CC prokaryotic or eukaryotic host cells. They are useful in assays to
CC identify putative ligands, which may be receptor agonists or
CC antagonists or bind with little effect, and which may be useful in
CC understanding obesity and for its prevention and treatment.
SQ Sequence 1162 AA;

alignment_scores:

Quality: 61.50 Length: 111

Ratio: 0.992 Gaps: 8

Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:

US-09-030-606-223 x W23398 ..

Align seg 1/1 to: W23398 from: 1 to: 1162

```
17 AAAACAATCT.....TCATTCAGAAAAATATCTTA..... 49
|||||:||||| ||| |||:| :|||:|
667 LysAsnAspSerLeucysSerValArgArgTyrValVallySHisArgTh 683
50 .....GGGACT.....GATATGGTAATTAAGTCAATTGA 80
||||| |||:| |||:| |||:| |||:|
683 rAlaHisAsnGlyThrTyrSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTTCTKGGGGCATTTCTTACAT.....TGCTCTGA 115
|||:| ||| ||| |||:| :|||:|
700 ThrPheLeuTyrPalagIuSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGCTGTGCGCAAAATTTGTATTTTATTTGAGACTTCT 165
:|:|:| ||| ||| |||:| :|||:|
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTyrProM 733
166 TATCAAAAGTAATGCTGCCAAGAAGCTCTAAGGAATAGTAGTCTCC 215
:|:|:| ||| ||| |||:| :|||:|
733 etSerIysVal.....AsnAlaValGlnSerLeuSerAlaTyrPro 746
216 MTCACCTGTTTGAGAGTGCTCTATCTAAAGATTGATTCCTGGAATG 265
:|:|:| |||:| :|||:| |||:| |||:|
747 LeuSerSer...SerCysValIleLeu.SerTyrThrLeuSerProAsnA 762
266 ACAATTATATTTTAACCTTTGGTGGGGGAA 294
|| :||| ||| ||| |||
762 sPTyrSerLeuLeuTyrLeuValIleGlu 771
```

seq_name: A_Geneseq_36:W34257

seq_documentation_block:

ID W34257 standard; Protein; 1162 AA.

AC W34257;

DT 11-MAY-1998 (first entry)

DE Rat wild-type ob receptor.

KW Ob receptor; obesity; leptin; isoform; ligand; rat.

OS Rattus sp.

PN WO9742340-A1.

PD 13-NOV-1997.

PF 02-MAY-1997; U07521.

PR 24-MAY-1996; GB-010995.

PR 06-MAY-1996; US-016899.

PA (MERI) MERCK & CO INC.

PI Caskey CT, Hess JW, Hey P, Phillips MS;

DR N-PSDB; T93103.

DR N-PSDB; T93103.

PT New isoform(s) of the Ob (leptin) receptor - used for identifying
 PT specific binding ligands, potentially useful for study, prevention
 PT and treatment of obesity
 PS Disclosure; Fig 1; 34pp; English.
 CC This protein sequence comprises wild-type (lean) rat ob receptor.
 CC The ob receptor has numerous isoforms resulting from alternative
 CC splicing; 3 novel isoforms, designated f, g and c' are disclosed
 CC (see W34258-60). These isoforms also apply to mouse and human ob
 CC receptors. The isoforms can be expressed in host cells,
 CC particularly E. coli, yeast or mammalian cells. They are used to
 CC identify specific binding ligands. Agonists, antagonists and
 CC ligand mimetics can be identified that are potentially useful in
 CC the study, prevention and treatment of obesity.
 SQ Sequence 1162 AA;

alignment_scores:
 Quality: 61.50 Length: 111
 Ratio: 0.992 Gaps: 8
 Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:
 US-09-030-606-223 x W34257 ..

Align seg 1/1 to: W34257 from: 1 to: 1162

```

17 AAAACAATCT.....TCATCAGAAAAATTAATCTA..... 49
|||||:||||| ||| |||:| :|||:
667 LysAsnAspSerLeucysSerValArggTyrrValValLysHisArgTh 683
50 .....GGGACT.....GATATGGTAATTATGGTCAATTTA 80
||||| |||:| |||:| |||:| |||:| |||:| |||:|
683 rAlaHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTRTTKGGGGCATTTCTTACAT.....TGCTTGA 115
||||:| ||| ||| |||:| |||:| |||:| |||:|
700 ThrPheLeuTrpAlaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGCTGTGCGCAAAATTTGCTATTTTATTTGGAGACTTCT 165
:::| |||:| |||:| |||:| |||:| |||:| |||:|
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTrpProM 733
166 TATCAAAAGTAATGCTGCCAAGAAGCTCTAAGCAATTAGTGTCC 215
::| ||| |||:| |||:| |||:| |||:| |||:| |||:|
733 etSerLysVal.....AsnAlaValGlnSerLeuSerAlaTyrrPro 746
216 MTCACCTGTTGGAGTGTGCTATTTCTAAAGATTTTGATTTCCCTGAATG 265
:::| |||:| |||:| |||:| |||:| |||:| |||:|
747 LeuSerSer..SerCysValIleLeu.SerTrpThrLeuSerProAsnA 762
266 ACAATTATATTTTAACTTTGGTGGGGAA 294
||| :||| ||| ||| |||
762 sPTyrSerLeuLeuTyrrLeuValIleGlu 771

```

seq_name: A_Geneseq_36:W75958

seq_documentation_block:

ID W75958 standard; Protein; 200 AA.
 AC W75958;
 DT 11-DEC-1998 (first entry)
 DE Mouse cell surface protein.
 KW Mouse; cell surface protein; thymocyte; lymphocyte; cell adhesion;
 KW signal transmission; autoimmune disorder; allergy; diagnosis;
 KW mitogen-stimulated.
 OS Mus sp.
 PN WO9838216-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1998; J00837.
 PR 26-FEB-1998; JP-062217.
 PR 27-FEB-1997; JP-062290.
 PA (NISB) JAPAN TOBACCO INC.
 PI Tamatani T, Tezuka K;
 DR WPI; 98=481144/41.

DR N-PSDB; V53200.
 PT Cell surface molecule expressed in thymocytes and lymphocytes and -
 PT mediating signal transmission and cell adhesion, and antibodies to
 PT it useful in treatment of auto-immune and allergic disorders.
 PS Claim 9; Page 110-112; 149pp; Japanese.
 CC The present sequence represents a mouse cell surface protein which is
 CC expressed by thymocytes and by mitogen-stimulated lymphocytes. The cell
 CC surface protein induces adhesion of mitogen-stimulated lymphocytes to
 CC antibodies recognising the cell surface protein. These antibodies also
 CC produce an increase in peripheral blood lymphocytes in the presence of
 CC an antibody recognising CD3 antigen. The cell surface protein contains
 CC the amino acid sequence FDDPPF in its extracellular region and the
 CC sequence YMFV in its intracellular region. The cell surface protein can
 CC be used in the prevention and treatment of autoimmune and allergic
 CC diseases, and in the diagnosis and investigation of such disorders.
 SQ Sequence 200 AA;

alignment_scores:
 Quality: 57.50 Length: 73
 Ratio: 1.513 Gaps: 2
 Percent Similarity: 52.055 Percent Identity: 32.877

alignment_block:
 US-09-030-606-223 x W75958 ..

Align seg 1/1 to: W75958 from: 1 to: 200

```

151 TATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAAGCTTAAGGA 200
|||||:| :|||:| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 TyrLeuHisIleTyrrGlnSerGlnLeucysCysGlnLeuLysLeuTrpLe 144
201 ATTAGTAGTGTCCCMTCACCTTGTGAGTGTGCTATTTCTAAAGATTT 250
||| :|||:| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 uProValGlyLeuProAlaPheValValValLeuLeupheGlyCysIleL 161
251 TGATTTCCTGCAATGACAATTATTTTAACCTTTGGTGGGGGAANAGTT 300
||||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 euIleIleTrp.....PheSerLysLysLysTyrrGlySerSerVal 174
301 ATAGGACCACAGCTCTCAGTCTGTAATTAATCTTTTATTCG 350
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
175 HisAspProAsnSerGlnTyrrMet.....PheMetAl 185
351 ACTGTGTTTGACCATTAAG 369
| ||| ||| ||| |||
185 aAlaValAsnThrAsnLys 191

```

seq_name: A_Geneseq_36:R58854

seq_documentation_block:

ID R58854 standard; Protein; 383 AA.
 AC R58854;
 DT 13-APR-1995 (first entry)
 DE Fowlpox virus (FPV) genomic EcoRI fragment on vector 443-88.8.
 KW Homology vector 443-88.8.
 OS Fowlpox virus.
 PN WO9419014-A.
 PD 01-SEP-1994.
 PF 28-FEB-1994; U01826.
 PR 26-FEB-1993; US-024156.
 PA (JAPG) NIPPON ZEON KK.
 PA (SYTR) SYNTRO CORP.
 PI Cochran MD;
 DR WPI; 94-294007/36.
 DR N-PSDB; Q68935.
 PT New recombinant fowl pox virus for use in vaccines - contains
 PT genes expressing antigens of Newcastle disease virus and opt.
 PT infectious bronchitis virus
 PS Example; Page 54-55; 85pp; English.
 CC The homology vector 443-88.8 contains a 3.5 kb FPV genomic
 CC EcoRI fragment which maps to the approx. 5.5 kb overlap of
 CC FPV genomic fragments Sali C and PstI F. The ORF spans a

CC HpaI insertion site. This ORF shows no AA sequence homology
CC to any known pox virus genes.
SQ Sequence 383 AA;

alignment_scores: Length: 100
Quality: 58.00
Ratio: 1.415 Gaps: 7
Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:

US-09-030-606-223 x R58854 ..

Align seg 1/1 to: R58854 from: 1 to: 383

```
99 TCCTTACATGTCCTGACAGATTAA...ATGCTGTGCCAAATTG 145
|||||
43 SerleuHiscysPheAspArgSerLysGlyLeuAspGlnProLysThrPh 59
146 TATTTTATTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGACT 195
|||||
59 eileLeuProGlyLysTyrSerAsn..... 67
196 AAGGAATTAGTAGTTCCTCCMTCACTGTTGGAGTGTCTATT..... 239
:::
68 .....AsnSerileLys.....LeuGluValAlaIleAspThr 78
240 ...CTAAAGATTGATTTCCTGGAATGACAAATTATTTTAACTTTGG 286
|||||
79 TyrLysLysAspSerAspPhe..... 85
287 TGGGGGAAANAGTTATAGACACACAGCTTCTGATCTGTAAT 336
|||||
86 .....SerTyr.....SerHisProCysGlnI 93
337 TAATCTTTATGCACTTGT.....TTGACCAT 365
:::
93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheAspHis 109
```

seq_name: A_Geneseq_36:R58600

seq_documentation_block:

ID R58600 standard; Protein; 383 AA.
AC R58600;
DT 14-APR-1995 (first entry)
DE Fowlpox virus protein encoded by homology vector 443-88.8.
KW Homology vector 443-88.8; recombinant fowlpox virus.
OS Fowlpox virus.
PN WO9419015-A.
PD 01-SEP-1994.
PF 28-FEB-1994; U02252.
PR 26-FEB-1993; US-024156.
PA (SYTR) SYNPRO CORP.
PI Cochran MD;
DR WPI: 94-294008/36.
DR N-PSDB: Q70574.
PT New recombinant fowl pox viruses - useful as vaccines against
PT fowl pox virus, Newcastle Disease Virus and infectious
PT laryngotracheitis virus.
PS Example; Page 63-64; 97pp; English.
CC Homology vector 443-88.8 contains a 3.5 kb fowlpox virus genomic
CC EcoRI fragment and is useful for the insertion of foreign DNA into
CC FPV. This EcoRI fragment maps to the approx. 5.5 kb overlap of FPV
CC genomic fragments SalI C and PstI F. There is a unique HpaI site
CC which is designated the 680 site. This site interrupts the ORF
CC in Q70574 at AA 226. This ORF shows no AA sequence homology to any
CC known pox virus genes.
SQ Sequence 383 AA;

alignment_scores: Length: 100
Quality: 58.00
Ratio: 1.415 Gaps: 7

Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:

US-09-030-606-223 x R58600 ..

Align seg 1/1 to: R58600 from: 1 to: 383

```
99 TCCTTACATGTCCTGACAGATTAA...ATGCTGTGCCAAATTG 145
|||||
43 SerleuHiscysPheAspArgSerLysGlyLeuAspGlnProLysThrPh 59
146 TATTTTATTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGACT 195
|||||
59 eileLeuProGlyLysTyrSerAsn..... 67
196 AAGGAATTAGTAGTTCCTCCMTCACTGTTGGAGTGTCTATT..... 239
:::
68 .....AsnSerileLys.....LeuGluValAlaIleAspThr 78
240 ...CTAAAGATTGATTTCCTGGAATGACAAATTATTTTAACTTTGG 286
|||||
79 TyrLysLysAspSerAspPhe..... 85
287 TGGGGGAAANAGTTATAGACACACAGCTTCTGATCTGTAAT 336
|||||
86 .....SerTyr.....SerHisProCysGlnI 93
337 TAATCTTTATGCACTTGT.....TTGACCAT 365
:::
93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheAspHis 109
```

seq_name: A_Geneseq_36:W10692

seq_documentation_block:

ID W10692 standard; Protein; 383 AA.
AC W10692;
DT 05-MAY-1997 (first entry)
DE Homology vector 443-88.8-encoded protein.
KW Fowlpox virus; FPV; recombinant virus; vector; vaccine;
KW Immunisation; Newcastle disease virus; NDV; poultry.
OS Fowlpox virus.
PN WO9640880-A1.
PD 19-DEC-1996.
PF 04-JUN-1996; U11187.
PR 07-JUN-1995; US-484790.
PA (SYTR) SYNPRO CORP.
PI Cochran MD, Junker DE, Singer PA;
DR WPI: 97-087060/08.
DR N-PSDB: T48511.
PT New recombinant fowlpox virus - contg. a foreign DNA sequence
PT inserted into the fowlpox virus genome, used for the prodn. of
PT vaccines.
PS Example 1A; Page 94-95; 134pp; English.
CC A polypeptide (W10692) that shows no homology to any known pox
CC virus protein is encoded by an open reading frame found in a 1.5 bp
CC fragment (T48511) of vector 443-88.8. This vector contains a 3.5
CC kb fowlpox virus (FPV) genomic EcoRI fragment that maps to the
CC approx. 5.5 kb overlap of FPV genomic fragments SalI C and PstI F
CC and contains a unique HpaI site that interrupts the open reading
CC frame. Vector 443-8.8 is useful for the insertion of foreign DNA
CC into FPV. Insertion of an SfiI fragment (T48510) contg. Newcastle
CC disease virus (NDV) haemagglutinin (HN) and fusion (F) genes
CC yielded vector 502-26.22 (see also T48502-04), which can be used to
CC insert NDV HN and F genes into FPV. The recombinant virus is used
CC to deliver the vaccine antigens to poultry.
SQ Sequence 383 AA;

alignment_scores: Length: 100
Quality: 58.00
Ratio: 1.415 Gaps: 7
Percent Similarity: 41.000 Percent Identity: 28.000

252 roglmetserglnpnetrcysgluasphe TyrpheProTirpleu 268
161 CTCTATC 169
269 leuilelle 271

This Page Blank (uspto)

OM of: US-09-030-606-223 to: Issued_Patents_AA:* out_format : pfs
Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODE=frame+np.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09030606/runat.24091999_171616_29813/app-query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-223
Query length: 383
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

| Sequence | Strd Orig | ZScore | EScore | len | ! Documentation | .. | |
|---|-----------|--------|--------|-------|-----------------|-------|------|
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-01826A-4 | | | | 58.00 | 128.18 | 1.24 | 383 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-02252A-4 | | | | 58.00 | 128.18 | 1.24 | 383 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US95-14442A-109 | | | | 56.50 | 135.71 | 1.52 | 359 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-467-568-13 | | | | 53.50 | 116.50 | 5.93 | 359 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-748-485-4 | | | | 53.50 | 116.50 | 5.93 | 359 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-103-170-2 | | | | 53.50 | 116.50 | 5.93 | 359 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-050-319B-41 | | | | 51.00 | 131.76 | 7.71 | 39 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-982-41 | | | | 51.00 | 131.76 | 7.71 | 39 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-808-550-34 | | | | 51.50 | 115.46 | 10.58 | 230 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-5 | | | | 51.00 | 117.52 | 11.46 | 163 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-232-087A-9 | | | | 51.00 | 117.46 | 11.48 | 164 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-221-583-49 | | | | 49.00 | 135.79 | 11.94 | 15 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US95-04018-49 | | | | 49.00 | 135.79 | 11.94 | 15 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-424-641B-2 | | | | 51.00 | 111.99 | 13.37 | 284 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-162-081B-37 | | | | 52.00 | 101.53 | 13.58 | 1069 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-780-872-37 | | | | 52.00 | 101.53 | 13.58 | 1069 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-162-081B-36 | | | | 52.00 | 101.43 | 13.62 | 1080 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-780-872-36 | | | | 52.00 | 101.43 | 13.62 | 1080 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-309-512-9 | | | | 51.50 | 106.35 | 13.63 | 574 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US92-08756A-9 | | | | 51.50 | 106.35 | 13.63 | 574 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-468-847B-14 | | | | 51.00 | 109.96 | 14.14 | 348 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-167-628-2 | | | | 51.00 | 109.94 | 14.15 | 349 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-386-680-2 | | | | 51.00 | 109.94 | 14.15 | 349 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-459-717-2 | | | | 51.00 | 109.94 | 14.15 | 349 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-712-302-2 | | | | 51.00 | 109.94 | 14.15 | 349 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US96-08140-2 | | | | 51.00 | 109.94 | 14.15 | 349 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-087-732-2 | | | | 51.00 | 107.21 | 15.27 | 459 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-385-229-2 | | | | 51.00 | 107.16 | 15.29 | 461 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-243-010-1 | | | | 51.00 | 106.64 | 15.51 | 486 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-385-229-4 | | | | 51.00 | 106.00 | 15.79 | 518 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-596-291-3 | | | | 52.00 | 93.21 | 17.12 | 2465 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-09943-2 | | | | 52.00 | 93.20 | 17.12 | 2466 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-118-270-33 | | | | 50.00 | 107.55 | 19.91 | 337 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US93-08528-33 | | | | 50.00 | 107.55 | 19.91 | 337 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-468-847B-15 | | | | 50.00 | 107.23 | 20.09 | 348 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-900-565-1 | | | | 50.00 | 104.00 | 21.98 | 481 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-900-565-1 | | | | 49.50 | 107.99 | 22.57 | 281 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-959-943-7 | | | | 50.00 | 101.69 | 23.44 | 607 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-959-943-9 | | | | 50.00 | 101.32 | 23.68 | 630 |
| /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-782-298-2 | | | | 50.00 | 100.96 | 23.92 | 653 |
| /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-463-911-3 | | | | 49.00 | 110.78 | 23.96 | 185 |
| /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-07926-44 | | | | 48.50 | 113.97 | 25.15 | 117 |

/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-681-812-5 - 48.50 112.70 26.06 133 1
/cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-07926-30 - 48.50 112.40 26.28
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-358-171-2 + 49.00 106.27 27.16 291 1

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-01826A-4

seq_documentation_block:

Sequence 4, Application PC/TUS9401826A
GENERAL INFORMATION:
APPLICANT: Syntro Corporation, et al.
TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01826A
FILING DATE: 28-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01826A-4

alignment_scores:

| Quality: | 58.00 | Length: | 100 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.415 | Gaps: | 7 |
| Percent Similarity: | 41.000 | Percent Identity: | 28.000 |

alignment_block:

US-09-030-606-223 x PCT-US94-01826A-4 ..

Align seg 1/1 to: PCT-US94-01826A-4 from: 1 to: 383

| | | |
|-----|---|-----|
| 99 | TCCTTACATGTCCTGACAGATTAA..ATGCTGTCCTCAAAATTTTG | 145 |
| 43 | SerLeuHisCysPheAspArgSerIysGlyLeuAspGlnProIysThrPh | 59 |
| 146 | TATTTATTTGGAGACTTCCTTCAAAAGTAATGTCGCCAAAGAGTCT | 195 |
| 59 | eileLeuProGlyIysTyrSerAsn..... | 67 |
| 196 | AAGGAATTAGTAGTGTCCCTTCACCTGTTGGAGTGTCTATT..... | 239 |
| 68 |AsnSerIleLys.....LeuGluValAlaIleAspThr | 78 |
| 240 | ..CTAAAGATTGTTGATTTCTGGAATGACAAATTATTTTAACTTTGG | 286 |
| 79 | TyrIysLysAspSerAspPhe..... | 85 |
| 287 | TGGGGAAAGTAGTATAGGACACACAGCTTTCACCTTCTGATACTTTGAAAT | 336 |
| 86 |SerTyr.....SerHisProCysGlnI | 93 |


```

337 TAATCTTTATTCACCTGT.....TTTGACCAT 365
      :::::|||||
93  lephneglnphcycysalserglyasnphseserglylsargpheasphs 109
      :::::|||||

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep: PCT-US94-02252A-4

```

```

seq_documentation_block:
; Sequence 4, Application PC/TUS9402252A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02252A
; FILING DATE: 28-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-02252A-4

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 58.00 | Length: 100 |
| Ratio: | 1.415 | Gaps: 7 |
| Percent Similarity: | 41.000 | Percent Identity: 28.000 |

```
alignment_block:
US-09-030-606-223 x PCT-US94-02252A-4 . . .
```

Align seg 1/1 to: PCT-US94-02252A-4 from: 1 to: 383

```

99 TCCTTACATGTCTCTTGACAAGATTAA...ATGCTGTGCCAAATTTTG 145
   |||||
43 SerIeuHIScysPheaspargSerIysGlyLeuaspGlnProIysThrPh 59
   |||||
146 TATTTTATTTGAGACTCTTTATCAAAAGTAATGCTGCCAAAGAGTCT 195
   |||||
59 eIleuProGlyIysTyrSerasn..... 67

196 AAGGAATTAGTAGTGTGCCMTCACCTGTTGGAGTGTGCTATT..... 239
   :::::
68 .....AsnSerIleLys.....LeuGluValAlaIleaspThr 78
   |||||

240 ...CTAAAGATTTTGATTTCCGTGAATGACAATTATATTTTAACTTGG 286
   |||||
79 TyrIysLysaspSerasphe..... 85

287 TGGGGGAANAGTTATAGGACCACAGTCTTCACTTCTGTGACTTGTAAAT 336
   |||||
86 .....SerTyr.....SerHisProCysGlnI 93

```

```

337 TAATCTTTATTGCACTGT.....TTGACCAT 365
      :::::||||:::|||||||
93 IephegIpnhecysvalsergIyasnphesergIyLysargphea$phIs 109
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pcp:PCT-US95-14442A-109

```

```

seq_documentation_block:
; Sequence 109, Application PC/TUS9514442A
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Arfsten, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFECTION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-14442A-109

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 56.50 | Length: 43 |
| Ratio: | 2.093 | Gaps: 1 |
| Percent Similarity: | 62.791 | Percent Identity: 32.558 |

alignment_block:
US-09-030-606-223/rev x PCT-US95-14442A-109 .

Align seg 1/1 to: PCT-US95-14442A-109 from: 1 to: 119

232 CACTCCAAACAAGTGAAGGGAACAAGTACTAATTCCTTAGACTTCCTTTGG 183
 |||:: |||:: ||| ::| |||:::|
 55 HlSGlnThrGlnIleLysGlyGlyIuIleCysLysAlaThrGlyTrpGI 71
 182 CAGCATTACTTTGATAGAAGTCTCCAATAAATACAAATTTTGCCA 133
 |::: |||:: ||| ||| |||:::|
 71 yargLengIyAlaAspGlnProValProAsnLysLengInLengIuT 88

132 CAGACATTTTAATCTGTCAAGACAATGT 104
|| ||| ||| :|||
88 hr...lleAlaileSeraspGluLysCys 96

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-467-568-13

seq_documentation_block:

Sequence 13, Application US/08467568
Patent No. 5817477
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
TITLE OF INVENTION: ADRENERGIC RECEPTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,568
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-324
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-568-13

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 53.50 | Length: | 74 |
| Ratio: | 1.408 | Gaps: | 2 |
| Percent Similarity: | 51.351 | Percent Identity: | 28.378 |

alignment_block:

US-09-030-606-223/rev x US-08-467-568-13 ..

Align seg 1/1 to: US-08-467-568-13 from: 1 to: 359

368 TTAATGTCAAACAAGTGCATTAAGATTATTACAGATCAGAAG 319
:::||||| :||| :|||
99 ValMetLeuCysThrAlaSerIleLeuAsnLeuPheMetIleSerLeuAs 115
318 TGAAGACTGTGCT.....CCTATACTNTTTCCC 290
:::||||| :||| :|||
115 pArgTyrCysAlaValMetaspProLeuArgTyrProValLeuValThrP 132
289 CCACCAAGT.....AAATATAA 270
||:||||| :||| :|||
132 roValArgValAlaIleSerLeuValLeuIleTyrValIleSerIleThr 148
269 TTGTCATTCCAGGAATCAAAATCTTTAGATAGACACACTCCAAGAAG 220
||||||| :||| :|||
149 LeuSerPheLeuSerIleHisLeuGlyTyrPAsnSerArgAsnGluThrSe 165

219 TGAKGGAGACACTAATATCC 198
| ||||| ||| :|||
165 rLysGlyAsnHisThrThrSer 172

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-748-485-4

seq_documentation_block:

Sequence 4, Application US/08748485
Patent No. 5817480
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guegler, Karl J.
APPLICANT: Goll, Surya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: NOVEL HISTAMINE H2 RECEPTOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,485
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0159 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 184088
US-08-748-485-4

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 53.50 | Length: | 74 |
| Ratio: | 1.408 | Gaps: | 2 |
| Percent Similarity: | 51.351 | Percent Identity: | 28.378 |

alignment_block:

US-09-030-606-223/rev x US-08-748-485-4 ..

Align seg 1/1 to: US-08-748-485-4 from: 1 to: 359

368 TTAATGTCAAACAAGTGCATTAAGATTATTACAGATCAGAAG 319
:::||||| :||| :|||
99 ValMetLeuCysThrAlaSerIleLeuAsnLeuPheMetIleSerLeuAs 115
318 TGAAGACTGTGCT.....CCTATACTNTTTCCC 290
:::||||| :||| :|||
115 pArgTyrCysAlaValMetaspProLeuArgTyrProValLeuValThrP 132

||||||| |||:||||| |||:|
19 PROHISGlnIleCysAsnValValAlaIleProGlyAsnAlaSerMet 34

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-982-41

seq_documentation_block:

; Sequence 41, Application US/08465982
; Patent No. 5863786

; GENERAL INFORMATION:

; APPLICANT: M.Feldmann, P.W. Gray,
; APPLICANT: M.J.C. Turner, F.M. Brennan

; TITLE OF INVENTION: Modified human TNFalpha (Tumor
; TITLE OF INVENTION: Necrosis Factor alpha) Receptor

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Reed & Roblins
; STREET: 635 Bryant Street

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/465,982

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/050,319

; FILING DATE: 10-May-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Robblins, Roberta L.

; REGISTRATION NUMBER: 33,208

; REFERENCE/DOCKET NUMBER: 5150-0030

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 617-8999

; TELEFAX: (415) 327-3231

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 39 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-465-982-41

alignment_scores:

Quality: 51.00 Length: 16

Ratio: 3.923 Gaps: 0

Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:

US-09-030-606-223/rev x US-08-465-982-41 ..

Align seg 1/1 to: US-08-465-982-41 from: 1 to: 39

291 CCCACCAAGTTAAATATTAATGTCATTCAGGAATCAAAATCTT 244

||||||| |||:||||| |||:|

19 PROHISGlnIleCysAsnValValAlaIleProGlyAsnAlaSerMet 34

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-808-550-34

seq_documentation_block:

; Sequence 34, Application US/08808550
; Patent No. 5871992

; GENERAL INFORMATION:

; APPLICANT: Teebor, George W.

; APPLICANT: Hilbert, Timothy P.

; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND

; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/808,550

; FILING DATE: 26-FEB-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 1049-1-001 N

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 230 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: S. cerevisiae

; STRAIN: Sce nfe-S

US-08-808-550-34

alignment_scores:

Quality: 51.50 Length: 43

Ratio: 1.776 Gaps: 2

Percent Similarity: 67.442 Percent Identity: 37.209

alignment_block:

US-09-030-606-223 x US-08-808-550-34 ..

Align seg 1/1 to: US-08-808-550-34 from: 1 to: 230

168 TCAAAAGTATGCTGCCAAGAGCTAAGGAATTAGTGTCCCMT 217

|||||:||||| |||:||||| |||:|

5 SerLysIleValaProValaSpIleIleGlyGlySerSerIleProva 21

218 CACT..TGTTGAGTGTGCTATTCTAAAGATTTGATTTCCGGAAT 264

|||:||||| |||:||||| |||:|

21 lThrValaIaSerLysCysGlyIleSerLysGlu.GlnIleSerProarg 37

265 GACAATTATATTTAACCTTGTTGGGG 291

|||:||||| |||:||||| |||:|

38 AsPTyrArgLeuGlnValLeuLeuGly 46

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-5

seq_documentation_block:

; Sequence 5, Application US/08219237B
; Patent No. 5874546

; GENERAL INFORMATION:

; APPLICANT: NAGATA, Shigekazu

; APPLICANT: ITOH, Naoto

; APPLICANT: YONEHARA, Shin

; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: James W. Hellwege
STREET: P.O. Box 2266 Eads Station
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,237B
FILING DATE: 28-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,129
FILING DATE: 22-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: James W. Hellwege
REGISTRATION NUMBER: 28,808
REFERENCE/DOCKET NUMBER: 516762
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-237B-5

alignment_scores:
Quality: 51.00 Length: 16
Ratio: 3.923 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:

US-09-030-606-223/rev x US-08-219-237B-5 ..

Align seg 1/1 to: US-08-219-237B-5 from: 1 to: 163

291 CCCACCAAGTTAAATATATATGTCATTCAGGAATCAAAATCTT 244
||||||| |||:~::~:~|||||||~::~:~
143 ProHISGlnIleCYsAsnValAlaIleProGlyAsnAlaSerMet 158

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-232-087A-9

seq_documentation_block:

Sequence 9, Application US/08232087A
Patent No. 5866372

GENERAL INFORMATION:

APPLICANT: Stein, Harold
APPLICANT: D rkop, Horst

APPLICANT: Latza, Ute
TITLE OF INVENTION: Lymphoid CD30-Antigen

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church
STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,087A

FILING DATE: 08-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 756-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 164 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
FEATURE:
NAME/KEY: Protein
LOCATION: 1..164
OTHER INFORMATION: /note= "TNFR2, see Fig. 5"
US-08-232-087A-9

alignment_scores:
Quality: 51.00 Length: 16
Ratio: 3.923 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:

US-09-030-606-223/rev x US-08-232-087A-9 ..

Align seg 1/1 to: US-08-232-087A-9 from: 1 to: 164

291 CCCACCAAGTTAAATATATATGTCATTCAGGAATCAAAATCTT 244
||||||| |||:~::~:~|||||||~::~:~
144 ProHISGlnIleCYsAsnValAlaIleProGlyAsnAlaSerMet 159

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-221-583-49

seq_documentation_block:

Sequence 49, Application US/08221583
Patent No. 5486595

GENERAL INFORMATION:

APPLICANT: Heavner, George A.
TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors

NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5486595r1s
STREET: One Liberty Place 46th Floor

CITY: Philadelphia
STATE: Pennsylvania

COUNTRY: USA
ZIP: 19403

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25:mdctcm0d.
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/221,583
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: CCOR-0185
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-221-583-49

alignment_scores:
Quality: 49.00 Length: 13
Ratio: 4.455 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 61.538

alignment_block:

US-09-030-606-223/rev x US-08-221-583-49 ..

Align seg 1/1 to: US-08-221-583-49 from: 1 to: 15

291 CCCACCAAGTTAAATATATGTCATTCAGGAAT 253
||||||| |||:::|||||||
3 ProHsglnlleAlaAsnValAlaIleProGlyAsn 15

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US95-04018-49

seq_documentation_block:

Sequence 49, Application PC/TUS9504018

GENERAL INFORMATION:

APPLICANT: Heavner, George A.

APPLICANT: Kruszynski, Marian

APPLICANT: Mervic, Miljenko

APPLICANT: Weber, Robert W.

TITLE OF INVENTION: Tumor Necrosis Factor Inhibitors

NUMBER OF SEQUENCES: 76

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

ADDRESS: Norris

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04018

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: CCOR-0232

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US95-04018-49

alignment_scores:
Quality: 49.00 Length: 13
Ratio: 4.455 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 61.538

alignment_block:

US-09-030-606-223/rev x PCT-US95-04018-49 ..

Align seg 1/1 to: PCT-US95-04018-49 from: 1 to: 15

291 CCCACCAAGTTAAATATATGTCATTCAGGAAT 253
||||||| |||:::|||||||
3 ProHsglnlleAlaAsnValAlaIleProGlyAsn 15

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-424-641B-2

seq_documentation_block:

Sequence 2, Application US/08424641B

Patent No. 5824523

GENERAL INFORMATION:

APPLICANT: Sylvain Moineau, Shirley A.

APPLICANT: Walker, Ebenezer R. Vedamuthu,

APPLICANT: and Peter A. Vandenberg

TITLE OF INVENTION: Isolated DNA Encoding

TITLE OF INVENTION: Enzyme For Phage

TITLE OF INVENTION: Resistance

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ian C. McLeod

STREET: 2190 Commons Parkway

CITY: Okemos

STATE: Michigan

COUNTRY: USA

ZIP: 48864

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

MEDIUM TYPE: storage

COMPUTER: Acer

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/424,641B

FILING DATE: April 19, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/366,480

FILING DATE: December 30, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ian C. McLeod

REGISTRATION NUMBER: 20,931

REFERENCE/DOCKET NUMBER: MT 4.1-151

TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 347-4100

TELEFAX: (517) 347-4103

TELEX: No. 5824523e

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 284 Amino Acids

TYPE: Amino Acid

STRANDEDNESS: Single

TOPOLOGY: linear

MOLECULE TYPE: Peptide

US-08-424-641B-2

alignment_scores:
Quality: 51.00 Length: 44
Ratio: 1.759 Gaps: 2
Percent Similarity: 65.909 Percent Identity: 31.818

alignment_block:

US-09-030-606-223/rev x US-08-424-641B-2 ..

Align seg 1/1 to: US-08-424-641B-2 from: 1 to: 284

```
184 GGCAGCATTTACTTTTGATAGAGTCTCCAAATATA.....ATACAAA 141
      |||::: ||| :::: |||::: |||::: |||:::
46  GLyGLyAlaLeuPhepGluProAlaProGlnLysAlaValIleAsnAs 62
      ::::: |||::: |||::: |||::: |||:::
140 TTTTGGCAGACAGACATTTTAACTCTGTCAAGACATGTAAGAAATGCCCC 91
      :|||::: |||::: |||::: |||::: |||:::
62  pPheAsnSerGluLeuIleAsnCysTyrArgGlnMetLysAspAsnProG 79
      :|||::: |||::: |||::: |||::: |||:::
90  AMAAYAYWATTAATTGACCATAATTACCAAT 59
      ::: |||::: ||| :::: |||:::
79  LngLLeuIleGluLeu.....LeuThrAsn 87
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-162-081B-37

seq_documentation_block:

; Sequence 37, Application US/08162081B
; Patent No. 5824492

; GENERAL INFORMATION:

APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter
APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,
APPLICANT: Stefano; Gout, Ivan Tarasovitch
TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,
TITLE OF INVENTION: THEIR PREPARATION AND USE
NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/162,081B

FILING DATE: February 7, 1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB93/00761

FILING DATE: 13 April 1993

ATTORNEY/AGENT INFORMATION:

NAME: Pasqualini, Patricia A.

REGISTRATION NUMBER: 34,894

REFERENCE/DOCKET NUMBER: LUD 5256

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 1069 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-162-081B-37

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 52.00 | Length: | 76 |
| Ratio: | 1.209 | Gaps: | 2 |
| Percent Similarity: | 56.579 | Percent Identity: | 25.000 |

alignment_block:

US-09-030-606-223 x US-08-162-081B-37 ..

Align seg 1/1 to: US-08-162-081B-37 from: 1 to: 1069

```
6  AAACAACAACAAAAAACAAATCTTCATTCAGAAAAATTATCTTAGCGACT 55
      |||::: |||::: ::||| |||::: |||:::
227 LysLysThrArgSerMetLeuLeuSerSerGlnGlnLeuLysLeuCysVa 243
      ::::: |||::: |||::: |||::: |||:::
56  GATATTGCTAATATATGCTCAATTTATWRTTRTKTGCGCATTTCTTAC 105
      ::::: |||::: |||::: |||::: |||:::
243 lLeuGluTyrGlnGlyLysTyrIleLeuLysValCysGlyCysAspGluT 260
      ::: |||::: |||::: |||::: |||:::
106 ATGTCTTGACAAGATTAAATGCTCTGTGCCAAATTT.....TGT 146
      ::: |||::: |||::: |||::: |||:::
260 YrPheLeuGluLysTyrProLeuSerGlnTyrLysTyrIleArgSerCys 276
      :|||::: |||::: |||::: |||::: |||:::
147 ATTTATTGAGACTTCTTATCAAAAGTAATGCTGCCAAAGAGTCTA 196
      |||::: |||::: ::: ::::: |||:::
277 lLeuLeuGlyArgMetProAsnLeuMetLeuMetAlaLysGluSer.. 292
      |||::: |||::: |||::: |||::: |||:::
197 AGGAATTAGTAGTGTTCCTCCMTCACTTGT 224
      |||::: |||::: |||::: |||::: |||:::
293 .LeuTyrSerGlnLeuProMetAspCys 301
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:39:06 ; Search time 2928.69 Seconds
(without alignments)
415.906 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383
Sequence: 1 AAAACAACAACAACAAAAA.....ATTAGCTATATGTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vl: *
18: em_fun: *
19: em_htg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vl: *
34: gb_htg1: *
35: gb_htg2: *
36: gb_in1: *
37: gb_in2: *
38: em_ba1: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pr4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | |
|------------|-------|--------------------------|-----------------------------|
| Result No. | Score | Query Match Length DB ID | Description |
| c 1 | 378.4 | 98.8 197441 34 | CNS00005 AL049779 Homo sapi |

| | | | | | | | |
|---|----|------|------|--------|----|------------|--------------------|
| c | 2 | 45.6 | 11.9 | 149764 | 10 | HSAC002076 | AC002076 Human BAC |
| c | 3 | 45.2 | 11.8 | 207139 | 34 | CEY17G7 | Z96049 Caenorhabd1 |
| c | 4 | 45.2 | 11.8 | 143092 | 36 | CEY17G7B | AL023828 Caenorhab |
| c | 5 | 45 | 11.7 | 104308 | 8 | AC006228 | AC006228 Genomic s |
| c | 6 | 44.6 | 11.6 | 82044 | 34 | PFMAL1P2 | AL031745 Plasmodi |
| c | 7 | 44.6 | 11.6 | 441 | 36 | IPU06960 | U06960 Ichneumon p |
| c | 8 | 44.2 | 11.5 | 151798 | 9 | HS435C23 | Z92844 Human DNA s |
| c | 9 | 44 | 11.5 | 172007 | 42 | AC006376 | AC006376 Homo sapi |
| c | 10 | 43.8 | 11.4 | 110680 | 8 | AC006259 | AC006259 Arabidops |
| c | 11 | 43.8 | 11.4 | 36823 | 10 | HSU209G1 | Z68873 Human DNA s |
| c | 12 | 43.8 | 11.4 | 216991 | 34 | HS164F3 | AL035422 Homo sapi |
| c | 13 | 43.2 | 11.3 | 4284 | 36 | DDU20661 | U20661 Dictyostell |
| c | 14 | 42.8 | 11.2 | 96642 | 8 | AC002291 | AC002291 Arabidops |
| c | 15 | 42.8 | 11.2 | 92612 | 8 | ATAC003974 | AC003974 Arabidops |
| c | 16 | 42.8 | 11.2 | 274690 | 34 | CEY60A9 | AL022281 Caenorhab |
| c | 17 | 42.8 | 11.2 | 167846 | 34 | CEY66C5 | Z98874 Caenorhabd1 |
| c | 18 | 42.8 | 11.2 | 24598 | 36 | CET24C2 | Z68120 Caenorhabd1 |
| c | 19 | 42.8 | 11.2 | 12029 | 37 | AE001400 | AE001400 Plasmodi |
| c | 20 | 42.6 | 11.1 | 300172 | 35 | AC005308 | AC005308 Plasmodi |
| c | 21 | 42.4 | 11.1 | 143751 | 11 | AC004160 | AC004160 Homo sapi |
| c | 22 | 42.4 | 11.1 | 142573 | 34 | HSDJ753D5 | AL049693 Homo sapi |
| c | 23 | 42.4 | 11.1 | 929 | 36 | DDIDDCOF1 | D37980 Dictyostell |
| c | 24 | 42.4 | 11.1 | 3946 | 36 | DDU25143 | U25143 Dictyostell |
| c | 25 | 42 | 11.0 | 143299 | 10 | HS811H13 | AL023805 Human DNA |
| c | 26 | 42 | 11.0 | 199606 | 35 | AC004688 | AC004688 Plasmodi |
| c | 27 | 42 | 11.0 | 194410 | 35 | AC005140 | AC005140 Plasmodi |
| c | 28 | 41.8 | 10.9 | 89281 | 8 | ATAC006304 | AC006304 Arabidops |
| c | 29 | 41.8 | 10.9 | 97494 | 8 | T15F16 | AF076275 Arabidops |
| c | 30 | 41.8 | 10.9 | 99902 | 10 | HS59B16 | AL032822 Human DNA |
| c | 31 | 41.8 | 10.9 | 86829 | 36 | PFMAL3P5 | AL034556 Plasmodi |
| c | 32 | 41.6 | 10.9 | 74342 | 7 | AB020742 | AB020742 Arabidops |
| c | 33 | 41.4 | 10.8 | 107331 | 34 | CEY49A10 | Z93240 Caenorhabd1 |
| c | 34 | 41.4 | 10.8 | 312766 | 34 | PFMAL4P3 | AL035476 Plasmodi |
| c | 35 | 41.4 | 10.8 | 37750 | 36 | CEZC504 | Z50029 Caenorhabd1 |
| c | 36 | 41.2 | 10.8 | 12039 | 3 | BTU25810 | U25810 Bos taurus |
| c | 37 | 41.2 | 10.8 | 79574 | 8 | ATAC002340 | AC002340 Arabidops |
| c | 38 | 41.2 | 10.8 | 156909 | 9 | AB020863 | AB020863 Homo sapi |
| c | 39 | 41.2 | 10.8 | 146813 | 11 | AF121898 | AF121898 Homo sapi |
| c | 40 | 41 | 10.7 | 299081 | 34 | AC006892 | AC006892 Caenorhab |
| c | 41 | 41 | 10.7 | 40168 | 36 | CELR06B10 | AF040654 Caenorhab |
| c | 42 | 40.8 | 10.7 | 83450 | 7 | AB009054 | AB009054 Arabidops |
| c | 43 | 40.8 | 10.7 | 82594 | 8 | ATAC005311 | AC005311 Arabidops |
| c | 44 | 40.8 | 10.7 | 154848 | 11 | AC002992 | AC002992 Homo sapi |
| c | 45 | 40.8 | 10.7 | 160262 | 11 | AC005242 | AC005242 Homo sapi |

ALIGNMENTS

| | | | | | | |
|------------|------------|--|---------------|--------------------|-----------------------|-------------|
| RESULT 1 | CNS00005/c | CNS00005 | 197441 bp | DNA | HTG | 06-MAY-1999 |
| LOCUS | | Homo sapiens | chromosome 14 | clone bac R-1012A1 | from RPCI-11 library, | |
| DEFINITION | | WORKING DRAFT SEQUENCE, in ordered pieces. | | | | |
| ACCESSION | | AL049779 | | | | |
| NID | | 94760257 | | | | |
| VERSION | | AL049779.1 | GI:4760257 | | | |
| KEYWORDS | | HTG; HTGS_PHASE2. | | | | |
| SOURCE | | human. | | | | |
| ORGANISM | | Homo sapiens | | | | |
| REFERENCE | | 1 (bases 1 to 197441) | | | | |
| AUTHORS | | Genoscope. | | | | |
| TITLE | | Direct Submission | | | | |
| JOURNAL | | Submitted (05-MAY-1999) Genoscope - Centre national de sequence | | | | |
| COMMENT | | 2, rue Gaston Cremieux - BP 191 91006 EVRY cedex - FRANCE E-mail : | | | | |
| FEATURES | | sequence.tendergenoscope.cns.fr | | | | |
| | | * NOTE: This is a 'working draft' sequence. | | | | |
| | | * This sequence will be replaced | | | | |
| | | * by the finished sequence as soon as it is available and | | | | |
| | | * the accession number will be preserved. | | | | |
| | | Location/Qualifiers | | | | |

source 1. .197441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="bac R-1012A1 from RPCI-11 library"
BASE COUNT 57409 a 42215 c 42060 g 55757 t
ORIGIN

Query Match 98.8%; Score 378.4; DB 34; Length 197441;
Best Local Similarity 98.2%; Pred. No. 5.3e-61;
Matches 376; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
Db 50969 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 50910
QY 61 TGGTAATATGTCATATTATWRTTKTGGGGCATTTCCCTTACATTTGCTTGACAAGA 120
Db 50909 TGGTAATATGTCATATTATATATTTTGGGGCATTTCCCTTACATTTGCTTGACAAGA 50850
QY 121 TTAATAATGCTGTGCCAAAAATTTGTATTATTGAGAGACTTCTATCAAAAGTAATGC 180
Db 50849 TTAATAATGCTGTGCCAAAAATTTGTATTATTGAGAGACTTCTATCAAAAGTAATGC 50790
QY 181 TGGCAAAGGAAGCTAAGGAATTAGTAGTGTCCMTCACTGTTGGAGTGTCTATTTC 240
Db 50789 TGGCAAAGGAAGCTAAGGAATTAGTAGTGTCCCATCACTGTTGGAGTGTCTATTTC 50730
QY 241 TAAAGATTTTGTATTCCTGGAATGACAATTATATTAACTTTGGTGGGGAANAAGTT 300
Db 50729 TAAAGATTTTGTATTCCTGGAATGACAATTATATTAACTTTGGTGGGGAANAAGTT 50670
QY 301 ATAGGACCACAGCTCTTCACTTCTGATACTTGTAAATTAATCTTTATGCACTGTTTGG 360
Db 50669 ATAGGACCACAGCTCTTCACTTCTGATACTTGTAAATTAATCTTTATGCACTGTTTGG 50610
QY 361 ACCATTAAAGCTATATGTTAAAA 383
Db 50609 ACCATTAAAGCTATATGTTAGAA 50587

RESULT 2
HSAC002076/c HSAC002076 149764 bp DNA PRI 12-MAY-1997
LOCUS
DEFINITION Human BAC clone GS345D13 from 7q31-q32, complete sequence.
ACCESSION AC002076
NID 92078461
VERSION AC002076.1 GI:2078461
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149764)
Maggi, L.
TITLE The sequence of H. sapiens BAC clone GS345D13
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 149764)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1997)
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or send an E-mail to egreen@nhgri.nih.gov

Mapping information for this clone was also provided by Dr. John D. McPherson, Department of Genetics/Genome Sequencing Center, Washington University School of Medicine.

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelobAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is H_GS274A07, 200 bp overlap. Actual start of this clone is at base position 1 of H_GS345D13; actual end is at 19827 of H_GS274A07. This cosmid lies in an unanchored cluster of unknown orientation.

This clone contains STS SWS54055 (NID:g1916534), SWS52840 (NID:g1113614), and SWS51647 (NID:g1113169).

FEATURES

source

1. 149764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="GS345D13"
/clone_lib="GSBAC1"
/map="7q31-q32"
3138. .4124
/rpt_family="L1"
4162. .4237
/rpt_family="L1"
complement(4345. .4638)
/rpt_family="ALU"
4769. .5162
/note="match to human EST AA039321 (NID:g1515599)
2k39d04.s1"
complement(5119. .5551)
/note="match to human EST AA039422 (NID:g1515699)
2k39d04.r1"
6029. .6148
/rpt_family="L1"
6787. .7179
/note="match to human EST AA150894 (NID:g1722424)
2147c03.s1"
6788. .7175
/note="match to human EST AA135883 (NID:g1696857)
2122b10.s1"
6788. .7201
/note="match to human EST N33317 (NID:g1153716)
yy08a02.s1"
6802. .7102
/note="match to human EST H01232 (NID:g864165) yj31f03.s1"
6909. .7179
/note="match to human EST H42319 (NID:g918371) y009g04.s1".
complement(7074. .7211)


```

misc_feature      /note="match to human EST W34968 (NID:g1316887)
                  mc33e06.r1"
                  complement(7074. .7211)
                  /note="match to human EST AA030363 (NID:g1497501)
                  m11c04.r1"
                  complement(7074. .7211)
                  /note="match to human EST W45750 (NID:g1330484)
                  mc79g10.r1"
                  complement(7084. .11162)
                  /gene="WUGSC:GS345D13.3"
                  complement(join(7084. .7209,11067. .11162))
                  /gene="WUGSC:GS345D13.3"
                  /note="GS345D13.3; match to protein P50152 (PID:g1730223)"
                  /codon_start=1
                  /product="G-protein gamma-11 subunit"
                  /protein_id="AAB54051.1"
                  /db_xref="PID:g2078464"
                  /db_xref="GI:2078464"
                  /translation="MPALHIEDPEKEKLMKEVEQLRKEYKLRQVSKSEELKNYIT
                  EERSGEDPIYKGIPEDKNPFKEKSCVIS"
                  7993. .8035
repeat_region    /rpt_family="L1"
                  complement(8127. .8417)
repeat_region    /rpt_family="ALU"
                  complement(11064. .11181)
                  /note="match to human EST AA030363 (NID:g1497501)
                  m11c04.r1"
                  complement(11064. .11181)
                  /note="match to human EST W45750 (NID:g1330484)
                  mc79g10.r1"
                  complement(11064. .11181)
                  /note="match to human EST W34968 (NID:g1316887)
                  mc33e06.r1"
                  complement(11065. .11236)
                  /note="match to human EST H00850 (NID:g863783) YJ31f03.r1"
                  complement(11065. .11383)
                  /note="match to human EST AA151020 (NID:g1722531)
                  z147c03.r1"
                  complement(11122. .11519)
                  /note="match to human EST N42976 (NID:g1166720)
                  yy08a02.r1"
                  complement(11137. .11476)
                  /note="match to human EST AA135882 (NID:g1696856)
                  z122b10.r1"
                  complement(13202. .13382)
                  /rpt_family="ALU"
                  complement(13681. .13971)
                  /rpt_family="ALU"
                  complement(14025. .14051)
                  /rpt_family="L1"
                  14675. .14792
repeat_region    /rpt_family="ALU"
                  complement(15009. .15303)
                  /rpt_family="ALU"
                  15416. .15674
repeat_region    /rpt_family="ALU"
                  complement(18751. .19015)
                  /rpt_family="ALU"
                  19128. .19385
repeat_region    /rpt_family="ALU"
                  20516. .20535
repeat_region    /rpt_family="L1"
                  22229. .22478
misc_feature      /note="match to human EST AA018874 (NID:g1482474)
                  ze56g05.s1"
                  complement(22263. .22510)
                  /note="match to human EST AA058783 (NID:g1551593)
                  zf58f06.r1"
                  complement(22382. .26553)
                  /gene="WUGSC:GS345D13.2"
                  complement(join(22382. .22510,26458. .26553))
                  /gene="WUGSC:GS345D13.2"
                  /note="GS345D13.2; match to protein Q08447 (PID:g585181)"

```

| LOCUS | DEFINITION | SEQUENCE | IN UNORDERED | PIECES |
|----------------|--------------------------------|-------------------|--|--|
| CEY17G7 | 207139 bp DNA | HTG | 03-DEC-1998 | WORKING DRAFT |
| ACCESSION | NID | VERSION | KEYWORDS | SOURCE |
| ORGANISM | Caenorhabditis elegans | HTGS_PHASE1 | Caenorhabditis elegans | Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis. |
| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
| 1 | Smye, R. | Direct Submission | Submitted (03-DEC-1998) | Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or tw@nematode.wustl.edu |
| On Aug 4, 1998 | this sequence version replaced | g1:2546913. | Order of segments is not known; 800 n's separate segments. | IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be |

CDS
 join(11683. .11753,11836. .11928,11980. .12304,13242. .13462,14218. .14331,14622. .14673,18228. .18428,18624. .18714,19438. .19658)
 /gene="Y17G7B.2c"
 /note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST yk470e5.3 comes from this gene; cDNA EST yk470e5.5 comes from this gene"
 /codon_start=1
 /protein_id="CAI19469.1"
 /db_xref="PID:e1356169"
 /db_xref="PID:g3947617"
 /db_xref="GI:3947617"
 /translation="MRSSKGGRGROAPKTAPTTVCYCDGKRELGSVEVVCSTLKWFGHRCLEFHEHLSNGVPEMICYFTCKQCRPTAEDWKAKADLVQMCVTLATLSAERLKADKLSAEHVPEDFTYLSLKDIEIVPYMENWYMLTAIKQKEWHQNLAPTLKEKNI FVQHNDDDLFALAEKNLSLGLPLHEAVKLIGKRPIERENREPHRIELPIEGPKTRGASKRRHAEPVTGKQKRLADYSSTAAPNGVQIDIPESKDNRYLLTEVDPNVEIPFIPLAFLSYIEIFQGDVLGCLHLVPDKLQIPANLSEKYLPLVSHKGFNLISFKANFEFEVQESADIATKLVEMPGSYIEFFHNGKSCGKAYENIYAGAYPYSISIFKSATATMNLGPKFRNLPRGATGIHARADEQOHEQTLSDMLYLVSEVNLDPVRVKREDDVDIKKEIKOEI"
 join(11683. .11753,11836. .11928,11974. .12304,13242. .13462,14218. .14331,15865. .16127,17187. .17299,18228. .18428,18624. .18714,19438. .19658)
 /gene="Y17G7B.2b"
 /note="cDNA EST EMBL:D35727 comes from this gene"
 /codon_start=1
 /protein_id="CAI19468.1"
 /db_xref="PID:e1323261"
 /db_xref="PID:g3947616"
 /db_xref="GI:3947616"
 /translation="MRSSKGGRGROAPKTAPTTVCYCDGKRELGSVEVVCSTLKWFGHRCLEFHEHFSRNSNGVPEMICYFTCKQCRPTAEDWKAKADLVQMCVTLATLSAERLKADKLSAEHVPEDFTYLSLKDIEIVPYMENWYMLTAIKQKEWHQNLAPTLKEKNI FVQHNDDDLFALAEKNLSLGLPLHEAVKLIGKRPIERENREPHRIELPIEGPKTRGASKRRHAEPVTGKQKRLADYSSTAAPNGVQIDIPESKDNRYLLTEVDPNVEIPFIPLAFLSYIEIFQGDVLGCLHLVPDKLQIPANLSEKYLPLVSHKGFNLISFKANFEFEVQESADIATKLVEMPGSYIEFFHNGKSCGKAYENIYAGAYPYSISIFKSATATMNLGPKFRNLPRGATGIHARADEQOHEQTLSDMLYLVSEVNLDPVRVKREDDVDIKKEIKOEI"
 20244. .23044
 /gene="Y17G7B.3"
 join(20244. .20381,21383. .21553,22571. .23044)
 /gene="Y17G7B.3"
 /note="similar to Metallo-beta-lactamase superfamily"
 /codon_start=1
 /protein_id="CAI19450.1"
 /db_xref="PID:e1323263"
 /db_xref="PID:g3947598"
 /db_xref="GI:3947598"
 /translation="MYVKSLLRRADNFMYIVKSSSEARAALLVDLVNEDYKELADKENIDITAVLTTHHYDHCGGNEGFRROFPNMVLIGDSRIIPAMDRHVKHGDMAEFAGLQIKCLSTPCHTSGHICYHITNPADSTSPGVFTGDTLFIAGCGRFEGTAPQMDVALNEILKNLPVETQIFPGHEYTVANLKFACHVEPGENEKAQKLEWAQROIEOGGFTVPSVVAEKATNPFMVRRESEIEIKSIGTCAVVGMAKLREMNKF"
 24223. .29110
 /gene="Y17G7B.4"
 join(24223. .24435,25013. .25162,26304. .26685,28307. .28435,28869. .29110)
 /gene="Y17G7B.4"
 /note="cDNA EST EMBL:D73584 comes from this gene"
 /codon_start=1
 /protein_id="CAI19451.1"
 /db_xref="PID:e1323264"
 /db_xref="PID:g3947599"
 /db_xref="GI:3947599"
 /translation="MSTNEAASQEDIALAQGAVLVKSCOVPDGSIPIRGFDSTASGPDFSLAILLSYMSYTGQATHLAQAIQOVNOMLSLPTPLCDDEKLFYPYEGROKRSCTIFLGITSNLVTSGLREVLRYCQVRNLVDCIVTSAGIEEDLIKLPKPSYLGFTTM DGAKLRNSGNMRAQNVLIIPNDNYCAFEDWMLPILDECLVEQEEKHLNWKPSKLIQRLG ERIGDESSILYMAAKHRIPVFCPALTDGSLGDMLYFHSVSKSSPGLRVDIVEDVRHINT

gene
 CDS
 IAVKSEFTGSIILGGVVKHHNNANLNRNGADHTVYINTGQEFSDSDSGAQPDEAVS
 WGVKPSAGAVKVAHEATLVFPLVAETFAKHEGHKD"
 complement(32881..37816)
 /gene="Y17G7B.5"
 complement(join(32881..33253,34402..35645,36629..37493,
 37653..37816))
 /gene="Y17G7B.5"
 /note="similar to MCM2/3/5 family; cDNA EST EMBL: T02370
 comes from this gene; cDNA EST yk244e4.5 comes from this
 gene"
 /codon_start=1
 /protein_id="CAI19452.1"
 /db_xref="PID:e1323265"
 /db_xref="PID:g3947600"
 /db_xref="GI:3947600"
 /translation="MADRANNDVDVROPLPIADDADDVDGIDEMENNEDEPEDEE
 GENLFDDMDERDYREQPELDQYSESGMDASDVGLSVSARRAEREMARQDLDD
 ALMYEDGDESEVDTRRRGRGRGDADDSVPMEEEDIPVDLENI RGTIRDHVS
 DEAVAKEIEREFKNFLRSFHEPGNKOTKYIQMIKSMADNRESLEVSFTLSDNGEO
 NISYFLPEAPENEMLAIMDRATEVVMNMYPEYRVCNEIKVRISQLPVEDIRMLROY
 HLNMLIRTAGVVTIASGILPQLAVVKYDCVACGYLLGPFVQNDDEV RPTICPSQOK
 GPELNTVENTVYHNYORITMOESPKNVAAGRLPKSKVDILGDLCDCKPGEIEVTG
 VYTNFEDGSLNFKOGFPVFNLTILHANHITNKDKMASDQLTDEDIKAI RELSQDPNISQ
 RVFSSIAPSIYGHDDVKRAIALALFRGEAKNPGAKHRLRGDINVLLCGDPTAKSQFL

Query Match 11.8%; Score 45.2; DB 36; Length 143092;
 Best Local Similarity 49.5%; Pred. No. 2;
 Matches 140; Conservative 1; Mismatches 140; Indels 2; Gaps 1;

| QY | 97 | TTTCCTTACATTGCTTGACCAAGATTAAATGCTGTCGCCAAATTTGTATTATTG | 156 |
|----|-------|---|-------|
| Db | 17486 | TTTGCTTATATTGATTAAAAAACAATAAAAATGCTGAAAAAGTTAGATTAAAG | 17545 |
| QY | 157 | GAGACTTCTATCAAAAGTAATGCTGCCAAAGAGTCTAAGGAATTAGTAGTTC | 216 |
| Db | 17546 | GAGGCTTCTGATACCGAATATCAATGCGAAAAAATCGAAAAATTTCCCTGATTTATA | 17605 |
| QY | 217 | TCACCTGTTGGAGTGTCTATTCTAAAGATTTCCTCGAATGACAATTATATT | 276 |
| Db | 17606 | TTAATTTTAAATCCAAAAATCCATTCGTTTAAATTCA--AACGACAAATTTTGA | 17663 |
| QY | 277 | TTAATCTTGGTGGGGAANAAGTTATAGACACAGCTCTCACTTCGATCTGTAAT | 336 |
| Db | 17664 | CTGATTACCGTCCGACAGTGTGTGATTTACGAGATTATATTAAATTCATTTT | 17723 |
| QY | 337 | TAATCTTTATGCACTGTTTGACCATTAAGCTATATGTTT | 379 |
| Db | 17724 | AAATTTTCATACCAATTTTCTGATTTTCGCTTTT | 17766 |

RESULT 5
 AC006228/c
 LOCUS
 DEFINITION
 AC006228 104308 bp DNA PLN 11-APR-1999
 Genomic sequence for Arabidopsis thaliana BAC F5J5, complete
 sequence.
 AC006228
 AC006228 94580732
 NID
 VERSION
 AC006228.4 GI:4580732
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 REFERENCE
 1 (bases 1 to 104308)
 CHAO, Q., SHINN, P., DUNN, P., BUEHLER, E., KAHN, S., KIM, C., WALKER, M.,
 WILLIAMS, S., ALTAFI, H., ARAUJO, R., CONN, L., CONWAY, A. B.,
 GONZALEZ, A., HANSEN, N. F., HUIZAR, L., KREMENTSKAYA, I., LENZ, C.,
 LI, J., LIU, S., LUIROS, S., ROWLEY, D., SCHWARTZ, J., TORIUMI, M.,
 VYSOTSKAYA, V., YU, G., DAVIS, R. W., FEDERSPIEL, N. A., THEOLOGIS, A. and
 ECKER, J. R.
 Genomic sequence for Arabidopsis thaliana BAC F5J5

| REFERENCE | AUTHORS | TITLE | JOURNAL |
|---------------------------|--|-----------------|----------------------|
| JOURNAL | Unpublished (1999) | | |
| REFERENCE | 2 (bases 1 to 104308) | | |
| AUTHORS | Ecker, J.R. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (18-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA | | |
| REFERENCE | 3 (bases 1 to 104308) | | |
| AUTHORS | Ecker, J.R. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (11-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA | | |
| COMMENT | On Apr 11, 1999 this sequence version replaced gi:4389503. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .104308 | | |
| | /organism="Arabidopsis thaliana" | | |
| | /db_xref="taxon:3702" | | |
| | /chromosome="1" | | |
| BASE COUNT | 33546 a 18958 c 18485 g 33319 t | | |
| ORIGIN | | | |
| Query Match | 11.7%; | Score 45; | DB 8; Length 104308; |
| Best Local Similarity | 48.5%; | Pred. No. 2.4; | |
| Matches 114; Conservative | 3; | Mismatches 118; | Indels 0; Gaps 0; |
| QY 19 | AAACAATCTCTCATTCAGAAAATTAATCTTAGGGAGCTGATATGGTAATATGTCATAT 78 | | |
| Db 56700 | AAATAAATCCCTCAAGTTGAGACATTATTACAGCTGGCTGCCACATTTAATAATTAT 56641 | | |
| QY 79 | TAAATWRTKTKTGGGCGCATTCCTTACATTTGCTTGACACAGATTAAATGCTGTGCCAA 138 | | |
| Db 56640 | TTTGATTAATTAGAAAGTAAATCTTTAAATTAATATTGACATTTAACTATCTCCCAA 56581 | | |
| QY 139 | AAATTTGATTTTATTGAGAGCTCTTATCAAAAGTAAGTGCAGCCAAAGAGCTAAG 198 | | |
| Db 56580 | GACTCTCTACATTACTACAGATTAGTTACTAAATAAATCAAACTCCAAATATTACAT 56521 | | |
| QY 199 | GAAATAGTAGTGTCCMTCACTGTTTGGAGTGTGCTATTCATAAAGATTTTGA 253 | | |
| Db 56520 | TATTTAAATAATTACAATAATATCATTTTGGATATTTGCTTTTGGACGACTATAA 56466 | | |
| RESULT 6 | | | |
| PEMALIP2/c | | | |
| LOCUS | PEMALIP2 82044 bp DNA HTG 09-APR-1999 | | |
| DEFINITION | Plasmodium falciparum chromosome 1 strain 3D7, WORKING DRAFT | | |
| SEQUENCE | SEQUENCE, in unordered pieces. | | |
| ACCESSION | AL031745 | | |
| NID | 94493855 | | |
| VERSION | AL031745.3 GI:4493855 | | |
| KEYWORDS | HTG; HTGS_PHASE1. | | |
| SOURCE | malaria parasite P. falciparum. | | |
| ORGANISM | Plasmodium falciparum | | |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | |
| AUTHORS | 1 (bases 1 to 82044) | | |
| | Bowman, S., Churcher, C., Harris, D., Lawson, D., Quail, M. and Barrell, B. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (09-APR-1999) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK | | |
| COMMENT | On Mar 24, 1999 this sequence version replaced gi:4455716. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum . IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. | | |
| | Order of segments is not known; 800 n's separate segments. | | |

```
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

Location/Qualifiers

1. .82044
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"

BASE COUNT 27301 a 7762 c 6747 g 29034 t 11200 others
ORIGIN

Query Match 11.6%; Score 44.6; DB 34; Length 82044;
Best Local Similarity 46.7%; Pred. No. 3;
Matches 134; Conservative 2; Mismatches 151; Indels 0; Gaps 0;

QY 1 AAACAACAAACAAAAAACAATTCCTCATCAGAAAAATTACTTAGGACTGATAT 60
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46959 AAAAAAAAAAAAAAAAAAATAATATATATATATATATATATTTTATTTATTTATTTATAT 46900

QY 61 TGGTAATTATGTGCAATTAAATWRTTKTGCGGCATTCCTTACATTTGCTTGACAAGA 120
 || ||| | : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46899 ATTACATATATATTATTTTAAATATTATTTATATATATTTTATTTTGAATATTT 46840

QY 121 TTAAATGTCTGTGCCAAAATTTTGATTTTATTTTGAGACTTCTATCAAAGTAATGC 180
 || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46839 TTTTATGTATTTTACAAATATATATAATTTTATTCGTAAATATTTATATGCGTATATG 46780

QY 181 TGCCAAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTC 240
 - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46779 TTATATTCGTGAGTGAGCTAATAAAGTACGTAATAAACAAACGATGATGGAATAT 46720

QY 241 TAAAGATTTGATTTCCCTGGAATGACAATTATTTTAACTTTGGT 287
 || ||| | - - - - - ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46719 TATTAAGATGTTACACATGTTAATATATATATGTTTATATATGTT 46673

RESULT 7
IPU06960 IPU06960 441 bp DNA INV 18-FEB-1995
LOCUS
DEFINITION Ichneumon promissorius mitochondrion 16S rRNA gene, partial
sequence.
ACCESSION U06960
NID 9463865
VERSION U06960.1 GI:463865

SOURCE
ORGANISM Ichneumon promissorius.
Mitochondrion Ichneumon promissorius
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
Ichneumonidae; Ichneumoninae; Ichneumon.

REFERENCE
AUTHORS 1 (bases 1 to 441)
TITLE Dowton,M. and Austin,A.D.
Molecular phylogeny of the insect order Hymenoptera: apocritan
relationships
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (21), 9911-9915 (1994)

MEDLINE
REFERENCE 95024071
TITLE 2 (bases 1 to 441)
AUTHORS Dowton,M.
Direct Submission
Submitted (21-FEB-1994) Mark Dowton, Department of Crop Protection,
Waite Campus, University of Adelaide, Adelaide, South Australia
5064, Australia

JOURNAL
FEATURES
SOURCE location/Qualifiers
1. .441
/organism="Ichneumon promissorius"
/mitochondrion
/db_xref="taxon:32393"
/dev_stage="adult"
<1..>441
/note="large subunit ribosomal RNA"

| | | | | |
|------------|---------------------|------|------|-------|
| BASE COUNT | 172 a | 30 c | 61 g | 178 t |
| ORIGIN | /product="16S rRNA" | | | |

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 11.6% | Score 44.6; | DB 36; | Length 441; |
| Best Local Similarity | 52.08; | Pred. No. 11; | | |
| Matches 92; Conservative | 3; | Mismatches 82; | Indels 0; | Gaps 0; |

QY 10 AAACAAAAAACAACTTCTTCATTCAGAAAAATATCTTAGGGACTGATATTGGTAATTA 69
||| ||| | ||| | | ||||| | | | |
Db 217 AAAAAAACAAATAATATTTGGTTGGGAATTAATAATTAATTAATTTAATTTT 276
QY 70 TGGTCATTTAATWTRTNTKTGGGCATTTCTTACATTTGCTTGACAAGATTAAATGT 129
| | | | : : : | | | | | | | | | |
Db 277 TTAACATGATTTTGATTTATATGATCTGATTTTTTTGGGATTAATAAAGATTAATAATAC 336
QY 130 CTGTGCCAAAATTTGTAATTTTATTTGGAGACTTCTTATCAAAAAGTAATGCTGCCAA 186
|| | | | | ||| ||| ||| ||| | | | |
Db 337 CTTAGGGAATAACAGCGTAATTTTTTTAGAGAGATCATATTAATTAATAAAGTTTGGCA 393

| | | | | |
|------------|---|------------|-----|-------------|
| RESULT | 8 | | | |
| HS435C23 | | | | |
| LOCUS | HS435C23 | 151798 bp | DNA | PRI |
| DEFINITION | Human DNA sequence from PAC 435C23 on chromosome X. Contains ESTs | | | 06-JUL-1997 |
| ACCESSION | 292844 | | | |
| NID | 92245344 | | | |
| VERSION | 292844.1 | GI:2245344 | | |
| KEYWORDS | X. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 151798) |
| AUTHORS | Mistry, S. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (04-JUL-1997) Sanger Centre, Hinxton, Cambridgeshire |

COMMENT

On Jul 7, 1997 this sequence version replaced gi:1877196.
IMPORTANT: This sequence is not the entire insert of clone A35C23.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together

with a note of the overlapping clone name. Note that the variations
annotated may not be found in the sequence submission corresponding
to

the overlapping clone as we submit sequences with only a small
overlap

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrx/>

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 435C23 is at 1 in this sequence. The true left end of clone 154K9 is at 151695.

435C23 is from the library RPC13 constructed at the Roswell Park Cancer

Institute by the group of Pieter de Jong.
For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES

Location/Qualifiers

```

source      1. .151798
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="435C23"
/clone_lib="RPC13"
1. .243
/note="AluSg repeat: matches 59. .300 of consensus;
incomplete repeat"
repeat_region
2813. .3129
/note="MLT2E repeat: matches 2. .303 of consensus"
3132. .3495
/note="MLT1A1 repeat: matches 1. .365 of consensus"
3551. .3580
/note="15 copies of 2 mer 87 & conserved"
3581. .3628
/note="MLT2FB repeat: matches 367. .414 of consensus"
4296. .4782
/note="MLT1D repeat: matches 505. .4. of consensus"
6002. .6302
/note="AluSx repeat: matches 1. .302 of consensus"
13751. .13834
/note="13 copies of 28 mer 87 & conserved"
13960. .14044
/note="MIR2 repeat: matches 55. .140 of consensus"
15122. .15170
/note="MLT2G repeat: matches 284. .331 of consensus"
15231. .15442
/note="MER20 repeat: matches 1. .202 of consensus"
15525. .15699
/note="MLT2_internal repeat: matches 2. .189 of consensus"
16152. .16888
/note="MLT2_internal repeat: matches 224. .1070 of
consensus"
16889. .17282
/note="MLT1B repeat: matches 390. .1 of consensus"
17283. .17698
/note="MLT2_internal repeat: matches 1066. .1491 of
consensus"
17720. .18175
/note="MLT1C repeat: matches 3. .466 of consensus"
18184. .19002
/note="MLT2_internal repeat: matches 1507. .2322 of
consensus"
19087. .19487
/note="MLT1E repeat: matches 101. .524 of consensus"
20292. .20596
/note="AluY repeat: matches 1. .301 of consensus"
21031. .21157
/note="AluJo repeat: matches 1. .131 of consensus;
incomplete repeat"
21187. .21371
/note="AluJo repeat: matches 121. .302 of consensus;
incomplete repeat"
22904. .23203
/note="AluJb repeat: matches 302. .2 of consensus"
26817. .26868
/note="26 copies of 2 mer 81 & conserved"
27867. .28302
/note="MSTA repeat: matches 1. .426 of consensus"
28625. .28691
/note="MLT1F repeat: matches 473. .539 of consensus"
30261. .30310
/note="25 copies of 2 mer 92 & conserved"
30698. .30835
/note="MIR2 repeat: matches 142. .1 of consensus"
31485. .32524
/note="L1MA3 repeat: matches 1049. .3 of consensus"
32377. .35673
/note="L1 repeat: matches 5390. .2079 of consensus"
36155. .36809
/note="L1 repeat: matches 2003. .1359 of consensus"
37304. .37577

```


repeat_region 38181. .38243
/note="MER43 repeat: matches 210. .272 of consensus"
repeat_region 38489. .38627
/note="MER43 repeat: matches 99. .236 of consensus"
repeat_region 41519. .42002
/note="MLT1D repeat: matches 503. .3 of consensus"
repeat_region 43031. .43115
/note="MIR repeat: matches 71. .155 of consensus"
repeat_region 43180. .43358
/note="MIR repeat: matches 262. .63 of consensus"
repeat_region 43568. .43762
/note="MSTB repeat: matches 1. .195 of consensus"
repeat_region 43763. .44064
/note="AluJ repeat: matches 1. .301 of consensus"
repeat_region 44119. .44324
/note="AluJ repeat: matches 302. .95 of consensus;
incomplete repeat"
repeat_region 46516. .46817
/note="AluY repeat: matches 300. .1 of consensus"
repeat_region 47484. .48286
/note="L1MA5 repeat: matches 1047. .238 of consensus"
repeat_region 48306. .48361
/note="2 copies of 28 mer 93 & conserved"
repeat_region 48401. .48645
/note="L1MA3 repeat: matches 233. .1 of consensus"
repeat_region 48484. .48898
/note="L1 repeat: matches 5390. .4990 of consensus"
repeat_region 48897. .49787
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region 49640. .50086
/note="L1 repeat: matches 5390. .4944 of consensus"
repeat_region 50085. .51156
/note="L1 repeat: matches 3871. .4944 of consensus"
repeat_region 51163. .52381
/note="L1 repeat: matches 4998. .3765 of consensus"
repeat_region 52389. .52714
/note="L1PB2 repeat: matches 899. .550 of consensus"
repeat_region 52705. .54411
/note="L1 repeat: matches 3787. .2073 of consensus"
repeat_region 54704. .54964
/note="L1PA1 repeat: matches 650. .912 of consensus"
repeat_region 56528. .56656
/note="MER5B repeat: matches 1. .130 of consensus"
repeat_region 56528. .56583
/note="MER5A repeat: matches 189. .136 of consensus"
repeat_region 56672. .57300
/note="L1MC3 repeat: matches 2485. .1826 of consensus"
repeat_region 57421. .57473
/note="MER5B repeat: matches 123. .177 of consensus"
repeat_region 59418. .59449
/note="16 copies of 2 mer 84 & conserved"
repeat_region 60640. .60935
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 61816. .61925
/note="MER42C repeat: matches 1421. .1537 of consensus"
repeat_region 63454. .63618
/note="MER20 repeat: matches 45. .218 of consensus"
repeat_region 63715. .64005
/note="AluJ repeat: matches 295. .2 of consensus"
repeat_region 64446. .64604
/note="MLT1B repeat: matches 3. .168 of consensus"
repeat_region 64655. .64774
/note="MLT1A2 repeat: matches 255. .374 of consensus"
repeat_region 65273. .65304
/note="16 copies of 2 mer 91 & conserved"
repeat_region 65305. .66197
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region 66048. .70552
/note="L1 repeat: matches 5390. .887 of consensus"
repeat_region 70539. .71298
/note="L1 repeat: matches 772. .22 of consensus"
repeat_region 74341. .74814
/note="L1 repeat: matches 4921. .5390 of consensus"

repeat_region 74668. .75555
/note="L1PA13 repeat: matches 1. .898 of consensus"
repeat_region 76249. .76352
/note="MIR repeat: matches 191. .82 of consensus"
repeat_region 76424. .76451
Query Match 11.5%; Score 44.2; DB 9; Length 151798;
Best Local Similarity 58.8%; Pred. No. 3;
Matches 70; Conservative 3; Mismatches 46; Indels 0; Gaps 0;
QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
Db 26844 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 26903
QY 61 TGGTAATTATGCTCAATTTAATWRTTCTTGGGCAATTCCTTACATTTGCTTGACAAG 119
Db 26904 AGCATATTATTCACAACTGCTTGTCTGTAGGCAGTGTGCTGTCATTTCTAGCTAAG 26962
RESULT 9
AC006376/c 172007 bp DNA PRI 24-MAR-1999
LOCUS Homo sapiens clone NH0386L03, complete sequence.
DEFINITION AC006376
ACCESSION g4508126
NID AC006376.2 GI:4508126
VERSION HTG.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172007)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172007)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 172007)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 24, 1999 this sequence version replaced gi:4204339.
FEATURES
source location/Qualifiers
1. 172007
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0386L03"
BASE COUNT 58408 a 33431 c 31325 g 48843 t
ORIGIN
Query Match 11.5%; Score 44; DB 42; Length 172007;
Best Local Similarity 47.2%; Pred. No. 3.1;
Matches 116; Conservative 5; Mismatches 125; Indels 0; Gaps 0;
QY 40 AATTATCTTAGGAGCTGATATGGAATATGTCATTTAATWRTTCTTGGGCAATTT 99
Db 66279 AACTATCTCAGGAAATAACTGTAGCACTTACATTTCAAAATAATTGGAATTAC 66220
QY 100 CCTTACATTTGCTTGACAAGATTAATGCTGTGCCAAATTTTGATTTTATTTGGAG 159
Db 66219 AGTAATTTTATTTGGTCGAGTGTAGTCATTTGCTGTATAGCATAGCTTTGTATTTATG 66160
QY 160 ACTTCTTATCAAAAGTAATGCTGCCAAAGAGTCTAAGGAATTAGTAGTGTCCMTCA 219
Db 66159 ATAGACATTGAAGGCTAAATATGAATAATGTAATTTATGCACTCTTAATTAGTGCAAT 66100
/note="L1 repeat: matches 4921. .5390 of consensus"

| QY | 220 | CTTGTTTGAGAGTGTGCTATTCCTAAAGATTTCGATTCCTGGATGACCAATTATATTTTA | 279 |
|------------|--|---|-----------------|
| Db | 66099 | ATAATTTTACTTTCTCTTCGACGACATTTTCAGTTAGCCTTGATGACTGTAATCCATGT | 66040 |
| QY | 280 | ACTTTG | 285 |
| Db | 66039 | ACTTTG | 66034 |
| RESULT | 10 | | |
| LOCUS | AC006259 | | |
| DEFINITION | AC006259 | 110680 bp DNA | PLN 28-DEC-1998 |
| ACCESSION | AC006259 | Arabidopsis thaliana BAC F21J6 from chromosome V, containing KNAT3 and mapping near 60.5 cm, complete sequence. | |
| NID | 94063730 | | |
| VERSION | AC006259.1 | GI:4063730 | |
| KEYWORDS | HTG. | | |
| SOURCE | thale cress. | | |
| ORGANISM | Arabidopsis thaliana | | |
| REFERENCE | 1 | (bases 1 to 110680) | |
| AUTHORS | Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R., Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R., Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N., Parnell, L.D. and McComble, W.R. | | |
| TITLE | Arabidopsis thaliana BAC F21J6 from chromosome V, near 60.5 cm | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 | (bases 1 to 110680) | |
| AUTHORS | Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R., Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R., Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N., Parnell, L.D. and McComble, W.R. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (28-DEC-1998) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA | | |
| COMMENT | BAC F21J6 is assigned to YAC C1C2E3 and maps to near 60.5 cm on the Lister & Dean RI map. Position 1 of F21J6, the T7 end, is oriented toward the telomere and position 110680, the SP6 end, is oriented toward the centromere. For more information on the mapping, sequencing and annotation of F21J6, please see http://www.cshl.org/Arabweb/F21J6-titlepage.html . A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by Genscan (http://CCR-081.mt.edu/GENSCAN.html), MZEF (http://www.cshl.org/genefinder) and GRAIL (http://compbio.ornl.gov/tools/index.shtml) and with splice sites predicted by NetPlantGene (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html). Alternate exons not used in building the gene models are presented on the web pages associated with F21J6. Genes are numbered according to the scheme BAC.gene.number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (http://expasy.hcuge.ch/sprot/prosite.html) and Pfam (http://pfam.wustl.edu/) libraries. A description of these categories can be found at http://muntjac.mips.biochem.mpg.de/Arab1/ . Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc. | | |
| FEATURES | | | |
| SOURCE | source | | |
| | 1..110680 | | |
| | /organism="Arabidopsis thaliana" | | |

```

gene
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /chromosome="V"
    /map="near 60.5 cM"
    <1..717
    /gene="F21J6.1"
    /note="encodes putative P450; entire gene encoded on T1IH3, GenBank accession number AC005964; genomic copy of EST Z33677; genomic copy of EST Z33963; gene model last edited on 16 Dec 98"
    /evidence=not_experimental
misc_feature
    1..24515
    /note="overlap with T1IH3, GenBank accession number AC005964, from position 67404 to 91916"
    /note="Region: overlap with T1IH3"
    join(<1..11,103..717)
    /gene="F21J6.1"
    /note="functional catalog ID=98"
    /codon_start=3
    /evidence=not_experimental
    /product="putative P450"
    /protein_id="AAC98443.1"
    /db_xref="PID:g4063733"
    /db_xref="GI:4063733"
    /translation="ILFNINAGIDTSAQVMWVTYLLISNPRVMKKAQAEVREVIKNKDIIEDIEDIERLEYLKMVKETFRVLPLVPLIPRBAASKDVKIGGYDIPKKTWIHNIWAHRNPVWKDPPEAFIPERMDNQIDYKGLNFELLPFGSGRRMCPGIGNGMALVHLT LINLYRFDMKLPEGMEVEDVDLEESYGLVCPRKVPQLIPLVLTQWS"
    complement(1310..1353)
    /note="function-putative_microsatellite"
    /rpt_type=tandem
    /rpt_unit=AT
    4808..6391
    /gene="F21J6.2"
    /note="encodes putative P450; gene model last edited on 16 Dec 98"
    /evidence=not_experimental
    join(4808..5683,5777..6391)
    /gene="F21J6.2"
    /note="functional catalog ID=98"
    /codon_start=1
    /evidence=not_experimental
    /product="putative P450"
    /protein_id="AAC98444.1"
    /db_xref="PID:g4063734"
    /db_xref="GI:4063734"
    /translation="MSLWTIIIVAFVFSSMIIVRIIRKTKNLPGPPRLPIIGNLHQ LGSKRSMFKLSETYGPLMSLKFGSVSTVASPTEVKELKTFDYCCSRPNMTYP ARVTYNLKDLCFSPYSKYRWREVRKMTVELVYAKRVQSFOHKTRKEEVALVDFIKQA SLEKPVNLKKIKMLKSGSVCRAVAFGINLOGSKLENTYEVEVIOGTVELVGSFAADYF PVVGRIIDRTIGLSKCEKLFKAMDAFDQSIKHLEDEIKDDIIDLLMERGETT LGFEQIDRLHTKGILANILNAGIDTSAQVMWVTYLLISNPRVLKKAQAEVREVIKHK DDIEEDIEDIERLEYLKMVKETFRINPLVPLIPRBAASKDVKIGGYNTPKKTWIHNIWA HRNPVWKDPPEAFIPERMDNQIDYKGLNFELLPFGSGRRICPGIGMGMA LVHLT LINLYRFDMKLPEGMKVADVLDLESYGLVCPKRIPLQLIPLVLTQWT"
    23757..24702
    /gene="F21J6.6"
    /note="mRNA sequence based on a comparison to cDNA L39646"
    23757..24702
    /gene="F21J6.6"
    /note="encodes ZFP3 zinc finger protein; genomic copy of cDNA L39646"
    /evidence=experimental
    23757..23810
    /gene="F21J6.6"
    /note="5'-UTR sequence based on a comparison to cDNA L39646"
    23811..24518
    /gene="F21J6.6"
    /note="functional catalog ID=98"
    /codon_start=1
    /evidence=experimental
    /product="ZFP3 zinc finger protein"

```



```

repeat_region /note="match: 3' EST R79568 clone 145988"
12107. .12148
/note="L1 element fragment"
repeat_region 12233. .12367
/note="L1 element fragment"
repeat_region 12577. .12870
/partial
/note="Alu repeat: matches 308. .1 of consensus"
16866. .16892
/note="9 copies of 3 mer 96 & conserved"
repeat_region 17081. .17123
/note="MER41 element fragment"
17583. .17871
/partial
/note="Alu repeat: matches 308. .1 of consensus"
18120. .19739
/note="L1 element fragment"
repeat_region 20360. .20484
/note="MSTC element fragment"
21043. .21180
/note="L1 element fragment"
repeat_region 21181. .21474
/partial
/note="Alu repeat: matches 308. .1 of consensus"
21528. .21647
/note="L1 element fragment"
repeat_region 21660. .21848
/note="MLT1B element fragment"
21662. .21713
/note="MLT1A element fragment"
repeat_region 21849. .21876
/note="7 copies of 4 mer 96 & conserved; 14 copies of 2
mer 96 & conserved"
21849. .21878
/note="5 copies of 6 mer 93 & conserved"
repeat_region 21927. .22036
/note="MLT1B element fragment"
21966. .22053
/note="MLT1A element fragment"
repeat_region 23206. .23253
/note="MER7 element fragment"
23344. .23570
/note="MER7 element fragment"
repeat_region 23586. .23959
/note="MLT1B element fragment"
23588. .23692
/note="MLT1C element fragment"
repeat_region 23745. .23839
/note="MLT1C element fragment"
repeat_region 24903. .25023
/partial
/note="Alu repeat: matches 148. .10 of consensus"
26975. .27080
/partial
/note="Alu repeat: matches 306. .193 of consensus"
27432. .27712
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 28071. .28220
/note="5 copies of 30 mer 81 & conserved"
28072. .28203
/note="22 copies of 6 mer 83 & conserved; 33 copies of 4
mer 83 & conserved"
28072. .28201
/note="5 copies of 2 mer 83 & conserved"
repeat_region 28074. .28213
/note="5 copies of 28 mer 81 & conserved"
28266. .28335
/note="MLT2C2 element fragment"
28389. .28628
/note="MLT2C2 element fragment"
repeat_region 28872. .28921
/note="2 copies of 25 mer 96 & conserved"
29191. .29274

```

```

repeat_region      /note="3 copies of 28 mer 82 % conserved"
                   30879. .31173
                   /partial
repeat_region      /note="Alu repeat: matches 308. .1 of consensus"
                   32683. .32857
                   /partial
repeat_region      /note="Alu repeat: matches 308. .133 of consensus"
                   32860. .32906
                   /partial
repeat_region      /note="Alu repeat: matches 47. .1 of consensus"
                   34532. .35401
                   /note="L1 element fragment"
repeat_region      35582. .35988
                   /note="L1 element fragment"
repeat_region      36039. .36219
                   /note="L1 element fragment"
repeat_region      36267. .36563
                   /partial
BASE COUNT      10072 a 7172 c 8356 g 11223 t
ORIGIN

```

| Query Match | 11.4%; | Score 43.8; | DB 10; | Length 36823; |
|-----------------------|---|----------------|----------------|---------------|
| Best Local Similarity | 51.0%; | Pred. No. 5.1; | | |
| Matches 102; | Conservative | 0; | Mismatches 98; | Indels 0; |
| | | | | Gaps 0; |
| QY 113 | TGACAAGATTAAAAATGCTGTGCCAAAATTTGTATTTATTTGGAGACTTCTTATCAAA | 172 | | |
| | | | | |
| Db 31264 | TAAATCAGATTAGAGAACTAATTTCTTAAACACACAGAAATTCACAGTTAAGATCTTTATAGA | 31323 | | |
| QY 173 | AGTAATGCTGCCAAAGSAGTCTAAGGAATTAGTAGTGTCCMTCACCTGTTTGGAGTG | 232 | | |
| | | | | |
| Db 31324 | AATAATATAAAAAATAAAAAGTACAGAGATTAGAACAGTGCAGTAAGATGATTTAAGTT | 31383 | | |
| QY 233 | TGCTATTTCTAAAAGATTTTGATTTCTCTGGAATGACAAATTATTTTAACTTTGGTGGGG | 292 | | |
| | | | | |
| Db 31384 | TAAATTTTAAATTTAATGCTGTTTGTGTGAGATTAACAGTTATTTTAAATTTAGAGTATG | 31443 | | |
| QY 293 | AAANAGTTATAGGACCACAG | 312 | | |
| | | | | |
| Db 31444 | GGCCAGATGATGGGACATAG | 31463 | | |

```

RESULT 12
HS164F3
LOCUS
DEFINITION Homo sapiens chromosome X clone 164F3, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035422
NID 95051831
VERSION AL035422.11 GI:5051831
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 216991)
Wilson, S.
Direct Submission
Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 12, 1999 this sequence version replaced gi:5050941.
COMMENT

```

* NOTE: This is a 'working draft' sequence.
Length: 216991 bp.

| | | | | |
|-----------------------|-----------------|-----------------|-----------|--------------|
| Query Match | 11.38; | Score 43.2; | DB 36; | Length 4284; |
| Best Local Similarity | 45.88; | Pred. No. 11; | | |
| Matches 135; | Conservative 3; | Mismatches 157; | Indels 0; | Gaps 0 |

| | | | |
|----|------|---|------|
| OY | 65 | AATTATGGCTCAATTTAATWRTTKTGCGGCAATTCCTTACATGTCTTGACAAGATTAA | 124 |
| | | : : | |
| Db | 3297 | AATTGAAGGATTAATAATTTTTTTTTTTTTTTTTTTTTTTTTTTCTTAATCCTTTTAATTTTC | 3356 |
| OY | 125 | AATGCTGTGCCAAAATTTTGTATTTTAATTTGGAGACTTCTTAATCAAAGAATAGCTGCC | 184 |
| | | | |
| Db | 3357 | GGTGTTCCTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTATTTTAATAAATAATGCATCA | 3416 |
| OY | 185 | AAAGGAAGCTTAAGGAATTAGTAGTGTCCCMTCACCTGTGTTGGAGTGTCTATTCTAA | 244 |
| | | : | |
| Db | 3417 | TATGAAATAAAAAAATCCAAAATAATTTGAAAAATCTATTTCAAAATAGATTGGTTAT | 3476 |
| OY | 245 | AGATTTTGATTTCTCGCATGACAAATTAATTTTAACCTTTGGTGGGGAANAGTTATAG | 304 |
| | | | |
| Db | 3477 | TGATTATTTTCTGGAATAATAAATAATAAACAAAAAATAAAAAAAAAAAATCT | 3536 |
| OY | 305 | GACCACAGCTCTCACCTCTGATACCTGTAATTAATCTTTATGCACTGTTTT | 359 |
| | | | |
| Db | 3537 | CCCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTCATTTCCCTTATTTT | 3591 |

| | | | | |
|------------|---|------------|-----|-----------------|
| RESULT 14 | | | | |
| LOCUS | AC002291 | 96642 bp | DNA | PLN 15-JUL-1998 |
| DEFINITION | Arabidopsis thaliana chromosome I BAC F2K20 genomic sequence, complete sequence. | | | |
| ACCESSION | AC002291 | | | |
| NID | g2477521 | | | |
| VERSION | AC002291.1 | GI:2477521 | | |
| KEYWORDS | HTG. | | | |
| SOURCE | thale cress. | | | |
| ORGANISM | Arabidopsis thaliana | | | |
| | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. | | | |
| REFERENCE | 1 (bases 1 to 96642) | | | |
| AUTHORS | Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vysotskaya,V.S., Yu,G., Ecker,J., Theologis,A. and Davis,R.W. Unpublished (1997) | | | |
| JOURNAL | 2 (bases 1 to 96642) | | | |
| REFERENCE | Federspiel,N.A., Davis,R.W., Conway,A.B., Palm,C.J., Conway,A.R., Kurtz,D.B., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Shin,P., Sun,H., Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vysotskaya,V., Yu,G., Theologis,A. and Ecker,J. Direct Submission | | | |
| TITLE | Submitted (05-JUN-1997) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | | | |
| JOURNAL | 3 (bases 1 to 96642) | | | |
| REFERENCE | Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vysotskaya,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W. Direct Submission | | | |
| TITLE | Submitted (07-OCT-1997) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | | | |
| JOURNAL | 4 (bases 1 to 96642) | | | |
| REFERENCE | Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vysotskaya,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W. Direct Submission | | | |
| TITLE | Submitted (04-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA | | | |
| JOURNAL | | | | |

COMMENT

94304, USA
On Oct 7, 1997 this sequence version replaced gi|2251218.
e-mail for correspondence: arabesequence.stanford.edu
Genes with similarity to proteins in the databases are described as
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge,
<http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene
(S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

FEATURES

source

1. .96642
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F22K20"
1692. .3328
/gene="F22K20.1"
join(1692. .2051,2148. .2235,2316. .2448,2542. .3328)
/gene="F22K20.1"
/note="similar to "tub" protein gpiU82468|2072162;
location of EST gb|T45528 and gb|AA395363"
/codon_start=1
/protein_id="AAC00626.1"
/db_xref="PID:g2829918"
/db_xref="GI:2829918"
/translation="MSFRSIYRDVRSIGSLSRSEDFKLSSLNKEGKSRSQSDSH
EEQLVTIQTWPANLPPELLRDVIKRLESESVPARRHVACASVCRRMDCKEI
VQSPELSGKITPEVPSLQPGPRDAMQCIFIKRDKSNLYHLYLCLSPALVENGFLL
SAKRIRRTTYTEYVISMHADPTISRSNTYIGKIRNFGLTKFIITDTQPAYNSNIARA
VQPVGLSRREYSKRVSYPKVGSKYIAOVSYELNVLTGRPRMHCAMNSIPASSLAIE
GGTVPGOPDIIVPRSILDESRSITSSSRKITVDYSDNFSSARFSIDLGPLSEDQEV
VLEEGERNSPPLVLANKPPEKHQLOQCWLNFGRVTVASVKNFOLIANKQPOPOPO
PQPQPPLTPQPPSGQTDGPDKIIILQFGKVGMFTMDFRYPLSAFAIPAICLSSFD
TLACE"

gene

11246. .12370
/gene="F22K20.2"
11246. .12370
/gene="F22K20.2"
/note="Unknown protein; location of EST gb|AA394987"
/codon_start=1
/protein_id="AAC00632.1"
/db_xref="PID:g2829924"
/db_xref="GI:2829924"
/translation="MESSPVACLPPDSLHQIFSSLPFRDIMICRSVCKEFPNQLTSQC
FIEIISTRPLLNLALRPPIHHHSRHSGNGHATNPRIHYVDPEQNQWRFNLDFL
PFERSPOVAASSGLIYMGDSIDLAESSSLVACNPLTRQFKLPLGLSAMSHTGLVL
VDSVNRYMVLTETLAALYSGTVVANQWLKFSSNLPSPKPRSVLMSSSVRALCDVGSPPW
RSQWKLFSCKLTNLTITHTNVLCLEKHEMGDIFDIKRPRLLRNGDSKLIIMGGLKS
TFSLNDACSTLILRLDLESLEWEAGRMPLMEYRGFESSKFKVFGGGDHYFSAKR
MGKIAMDCWGQMRWIEGVAYDGLCRGFVFDARKLTLP"

gene

complement(14545. .15792)
/gene="F22K20.3"
/note="Highly similar to atext1, extensin gene gb|U43627"
complement(14671. .15792)
/gene="F22K20.3"
/note="Highly similar to extensin (atExt1) protein
gi|1167961"
/codon_start=1
/product="extensin"
/protein_id="AAC00630.1"
/db_xref="PID:g2829922"
/db_xref="GI:2829922"
/translation="MASFLVALFSLAFVSQTTANTFYSSPPPVPVKHSPPPVYKSP
PKHYSPPVYKSPPPVKKHSPDPVYKSPPPVYKSPPPVYKSPPPVYKSPPPVYKSP
KHSPPPVYKSPPPVKKHSPDPVYKSPPPVYKSPPPVYKSPPPVYKSPPPVYKSP
SPPPVYKSPPPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYK
PPPYKSPPPVYKSPPPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSP
PPPYKSPPPVYKSPPPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSPDPVYKSP

[illegible]

QIVDINHDEDESDEFARLAHRSSTPTRRPVHGSDSGAVDILSGDVYKPKQGNSSSQ
GVRKPPPEPHTSSSSSPVFDDASPOQSSEVIRNLPRPSRHNRQGFEEHHSS
SGSDSYEGQTRNLSLTSSSEPQKEEPEDLLFKDLVEFAFATRRSSKANNNRSL"
gene
complement(45417. .46193)
/gene="F22K20.8"
CDS complement(45417. .46193)
/gene="F22K20.8"
/note="Hypothetical protein"
/codon_start=1

Query Match 11.2%; Score 42.8; DB 8; Length 96642;
Best Local Similarity 46.0%; Pred. No. 6;
Matches 131; Conservative 3; Mismatches 151; Indels 0; Gaps 0;

QY 19 AACACATTCTTCATTCAGAAAAATTATCTTAGGACTGATATTGGTAATATGTCGAATT 78
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10849 AATAAAGTCATGAGTTAAGATTATTATTACAATGGCTGTCACTGACATTTAATATTATT 10790
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 79 TAATWRTTKTGGGCATTTCCCTTACATTCCTTGACAAGATTAAATATGCTGTGCCAA 138
| : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10789 TTTTATATATTAGAAGAAAATCTTCTAATTAAATATTGATATTTAACAATCTTCTCAA 10730
| | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 139 AATTTGTATTTTATTGGAGACTTCTTACAAAGTAATGCTGCCAAAGAGCTTAAG 198
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10729 ATCTTTCTACTTTAACACACAATTAAATTACTAAATAAACTTCCAATAATTATAT 10670
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 199 GAATTAGTAGTGTCCCMTCACCTTGTGGAGTGTGCTATTCTAAAGATTTGATTTCC 258
| | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db 10669 CATTTAATTACTACAAATAATATCGTTTTTGATATTGCTTTCCGCATGATTAATCATG 10610
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 259 TGGAATGACATTAATTTTAACCTTGTGGGGGGAANAGTTATA 303
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10609 AAGCCGCATAGATATTGATGACATTTAATTACTACAAATTACA 10565
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
ATAC003974/c

LOCUS ATAC003974 92612 bp DNA PLN 19-AUG-1998
DEFINITION Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence,
complete sequence.
AC003974
ACCESSION AC003974
NID g2914688
VERSION AC003974.1 GI:2914688
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE
AUTHORS 1 (bases 1 to 92612)
Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
Somerville,C.R. and Venter,J.C.
JOURNAL Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence
TITLE Unpublished
REFERENCE 2 (bases 1 to 92612)
Rounsley,S.D. and Lin,X.
JOURNAL Rounsley,S.D. and Lin,X.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 92612)
Rounsley,S.D.
JOURNAL Direct Submission
AUTHORS Submitted (27-FEB-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Feb 27, 1998 this sequence version replaced gi:2828768.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA

gene

complement(<12375. .>13376)

/gene="F24L7.4"

/note="predicted by genefinder and genscan"

complement(12375. .13376)

/gene="F24L7.4"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAC04482.1"

/db_xref="PID:g2914692"

/db_xref="GI:2914692"

/translation="MMKIAIATERTDLPRLVSTAIRASNSQNASDVERCFDYLRLK
GLNLSVKNLSHSHKVIILPLESLRDHENPKIRTEAHVLEFTSMKRTFYSSGONSSSTCKNP
NPLKLKVVVKAACSELSKKKNEQLSHGFAVLKAKKETEYFLGMKKNEDERSRYHETREMK
QIGDSKSFALMRTIEKKKNSPFIQGSKNPRSGAGETGVYIKILRPDPFCKSVSPRP
PLMKKHQPVKAFENPKTCLYLAKKNSPEMLELFEMAKKSADVANAKGFLAAKEASIC
VDLALIMKFSIIISTAIETRIEMKLERLTKHKDRKICNALALLHHWROTLIRNOQ"

complement(13364. .13410)

repeat_region

| | | | | |
|---------------------------|--------|-----------------|-----------|---------------|
| Query Match | 11.2%; | Score 42.8; | DB 8; | Length 92612; |
| Best Local Similarity | 46.0%; | Pred. No. 6.1; | | |
| Matches 131; Conservative | 3; | Mismatches 151; | Indels 0; | Gaps 0; |

OY 19 AACCAATCTTCATTCAGAAAAATTAICTTAGGGA CTGATATTGGTAATTATGCCAATT 78
| | | | | | | | | | | | | | | | | |
Db 53198 AAATAAATCCTGGAGTTGAGATTTATTACAGTGCCTACC ACTGGCATTTAAATTTGTT 53139

OY 79 TAAATWTRRTTKTG GGCATTTCCCTTACATGCTTGACAAGATTAAAATGTCTG CCAA 138
| : : | | | | | | | | | | | | | | | |
Db 53138 TTGGATTAATTAGAAAGAAAATCTCTAAATTAATATT TGACATTTAACAAATCTTCCCAA 53079

OY 139 AATTTGTATTTTATTTGAGACTTCTATCAAAAGTAATGCTGC CAAAGAAAGTCTAAG 198
| | | | | | | | | | | | | | | | | |
Db 53078 ATCTTTCTACCTTA ACTACACAAATTAATTAATAAATAA AACTCCAAAATATTTAATAT 53019

OY 199 GAATTAGTAGTGTCCCMTCAC TTGTTGGAGTGTGCTATTCTAAAGATTGATTTCC 258
| | | | | : | | | | | | | | | | | |
Db 53018 CATTTAATTA CTACAAAATATATCATTTTGTGATAITGCTTTTG ACATGATTAATCATC 52959

OY 259 TCGAATGACCAATTATATTTTAAC TTTGGTGGGGGAAANA GTTATA 303
| | | | | | | | | | | | | | | | | |
Db 52958 AAGCCGATAGATAITTTGATAGCATTTAATTA CTACAAAATTACA 52914

Search completed: September 28, 1999, 12:41:49
Job time: 6189 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:28:48 ; Search time 289.74 Seconds
(without alignments)
330.723 Million cell updates/sec

Title: US-09-030-606-223
Perfect score: 383
Sequence: 1 AAACAACAACAACAAAAA.....ATTAAGCTATATGTTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 31585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| 1 | 380 | 99.2 | 383 | 1 | V58693 | Prostate tumour sp |
| 2 | 380 | 99.2 | 383 | 1 | V61268 | CDNA sequence of p |
| 3 | 40 | 10.4 | 110000 | 1 | X20248_08 | Continuation (9 of |
| 4 | 39.8 | 10.4 | 110000 | 1 | V21209_03 | Continuation (4 of |
| 5 | 39.4 | 10.3 | 110000 | 1 | X20248_06 | Continuation (7 of |
| 6 | 39.4 | 10.3 | 111309 | 1 | X20250 | Borrelia burgdorfe |
| 7 | 38.6 | 10.1 | 2287 | 1 | V22738 | Babesia microti BM |
| 8 | 38 | 9.9 | 4093 | 1 | V74419 | Staphylococcus aur |
| 9 | 37.6 | 9.8 | 2636 | 1 | V01871 | Human brain specif |
| 10 | 37.2 | 9.7 | 685 | 1 | X19427 | Nicotiana tabacum |
| 11 | 36.8 | 9.6 | 110000 | 1 | V21209_07 | Continuation (8 of |
| 12 | 36 | 9.4 | 2418 | 1 | Q27886 | P. falciparum GBP13 |
| 13 | 36 | 9.4 | 3607 | 1 | T89990 | Cryptosporidium pa |
| 14 | 35.8 | 9.3 | 110000 | 1 | V21209_05 | Continuation (6 of |
| 15 | 35.6 | 9.3 | 9047 | 1 | X12949 | Enterococcus faeca |
| 16 | 35.2 | 9.2 | 731 | 1 | Q21036 | Region specific to |
| 17 | 35.2 | 9.2 | 5064 | 1 | V69110 | Neisseria meningit |
| 18 | 35.2 | 9.2 | 400 | 1 | V69123 | Neisseria meningit |
| 19 | 35 | 9.1 | 8371 | 1 | T66909 | Human neuropeptide |
| 20 | 35 | 9.1 | 110000 | 1 | V21209_00 | Methanococcus jann |
| 21 | 35 | 9.1 | 600 | 1 | V86185 | EST clone J635. Ne |
| 22 | 35 | 9.1 | 14752 | 1 | X20256 | Borrelia burgdorfe |
| 23 | 34.8 | 9.1 | 10182 | 1 | V74427 | Staphylococcus aur |
| 24 | 34.6 | 9.0 | 5534 | 1 | Q35988 | Tomato hsp80 genom |
| 25 | 34.6 | 9.0 | 2334 | 1 | V09695 | M. hyopneumoniae 2 |
| 26 | 34.6 | 9.0 | 18359 | 1 | X20255 | Borrelia burgdorfe |
| 27 | 34.6 | 9.0 | 9842 | 1 | X20259 | Borrelia burgdorfe |
| 28 | 34.4 | 9.0 | 110000 | 1 | T58840_4 | Continuation (5 of |
| 29 | 34.4 | 9.0 | 3198 | 1 | T92702 | Candida Carhol gen |
| 30 | 34.4 | 9.0 | 3198 | 1 | T92869 | Candida Carhol gen |
| 31 | 34.4 | 9.0 | 1158 | 1 | V24720 | H. pylori ORF 06p |
| 32 | 34.4 | 9.0 | 1146 | 1 | V25046 | H. pylori cytoplasm |
| 33 | 34.4 | 9.0 | 1350 | 1 | X14148 | H. pylori GHPO 422 |
| 34 | 34.4 | 9.0 | 110000 | 1 | X20248_05 | Continuation (6 of |
| 35 | 34.4 | 9.0 | 110000 | 1 | X20248_05 | Continuation (6 of |
| 36 | 34.4 | 9.0 | 116277 | 1 | X20249 | Borrelia burgdorfe |
| 37 | 34.4 | 8.9 | 2749 | 1 | T66652 | Housefly Hermes el |
| 38 | 34.2 | 8.9 | 3164 | 1 | V26159 | Swinepox virus Hin |
| 39 | 34.2 | 8.9 | 1762 | 1 | V33193 | Secreted protein C |
| 40 | 34.2 | 8.9 | 14752 | 1 | X20256 | Borrelia burgdorfe |
| 41 | 34 | 8.9 | 5852 | 1 | Q11710 | Dictyostelium plas |
| 42 | 34 | 8.9 | 2503 | 1 | Q53480 | PNPX30 xylanase CD |
| 43 | 34 | 8.9 | 19124 | 1 | T72882 | Plasmodium var-7 g |

ALIGNMENTS

c 44 34 8.9 5849 1 V33135 Plasmodium berghei
c 45 34 8.9 11802 1 V74381 Staphylococcus aur

RESULT 1
V58693
ID V58693 standard; CDNA; 383 BP.
AC V58693;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone P5095.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 134; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 383 BP; 123 A; 51 C; 63 G; 140 T;

Query Match 99.2%; Score 380; DB 1; Length 383;
Best Local Similarity 100.0%; Pred. No. 7.3e-77;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
DB 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
QY 61 TGGTAATATGTCATTTAATWTRRTKTGGGCAATTCCTTACATGCTTGACAAGA 120
DB 61 TGGTAATATGTCATTTAATWTRRTKTGGGCAATTCCTTACATGCTTGACAAGA 120
QY 121 TTAATATGTCGTGCCAAATTTTGTATTTATTTGGAGACTTCTATCAAAAGTAATGC 180
DB 121 TTAATATGTCGTGCCAAATTTTGTATTTATTTGGAGACTTCTATCAAAAGTAATGC 180
QY 181 TGCCAAGAGAGCTTAAGGAATAGTAGTGTCCCMCTCACTGTTGGAGTGTCTATTC 240
DB 181 TGCCAAGAGAGCTTAAGGAATAGTAGTGTCCCMCTCACTGTTGGAGTGTCTATTC 240
QY 241 TAAAGATTTGATTTCCGTGAATGACAATATATATTTAACTTTGGTGGGGAANAAGTT 300
DB 241 TAAAGATTTGATTTCCGTGAATGACAATATATATTTAACTTTGGTGGGGAANAAGTT 300
QY 301 ATAGACACAGCTCTTCACTTCGATACCTGTAAATTAATCTTTATTCGACTGTTTG 360
DB 301 ATAGACACAGCTCTTCACTTCGATACCTGTAAATTAATCTTTATTCGACTGTTTG 360
QY 361 ACCATTAAGCTATATGTTTAAA 383
DB 361 ACCATTAAGCTATATGTTTAAA 383

RESULT 2
ID V61268 standard; cDNA; 383 BP.
AC V61268;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P509S.
KW prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 124; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 383 BP; 123 A; 51 C; 63 G; 140 T;

Query Match 99.2%; Score 380; DB 1; Length 383;
Best Local Similarity 100.0%; Pred. No. 7.3e-77;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATAT 60
DB 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATAT 60
QY 61 TGGTATATATGTCATTTAATWRTTKTGGGGCATTTCCCTACATTTGCTTGACAAGA 120
DB 61 TGGTATATATGTCATTTAATWRTTKTGGGGCATTTCCCTACATTTGCTTGACAAGA 120
QY 121 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGGAGACTTCTTACAAAGTAATGC 180
DB 121 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGGAGACTTCTTACAAAGTAATGC 180
QY 181 TGCCAAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTC 240
DB 181 TGCCAAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTC 240
QY 241 TAAAGATTTTGAATTCCTGGAATGACAATATATTTTAACTTTGGTGGGGAANAAGTT 300
DB 241 TAAAGATTTTGAATTCCTGGAATGACAATATATTTTAACTTTGGTGGGGAANAAGTT 300
QY 301 ATAGACACACAGTCTCACTCTGTACTCTGTAATTAATCTTTTATTTGCACTGTTTG 360
DB 301 ATAGACACACAGTCTCACTCTGTACTCTGTAATTAATCTTTTATTTGCACTGTTTG 360
QY 361 ACCATTAAAGCTATATGTTTAAAA 383
DB 361 ACCATTAAAGCTATATGTTTAAAA 383

RESULT 3

X20248_08
Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

| WP | Fragment Name | Begin | End |
|----|---------------|--------|--------|
| WP | X20248_00 | 1 | 110000 |
| WP | X20248_01 | 100001 | 210000 |
| WP | X20248_02 | 200001 | 310000 |
| WP | X20248_03 | 300001 | 410000 |
| WP | X20248_04 | 400001 | 510000 |
| WP | X20248_05 | 500001 | 610000 |

| | | | |
|----|-----------|--------|--------|
| WP | X20248_06 | 600001 | 710000 |
| WP | X20248_07 | 700001 | 810000 |
| WP | X20248_08 | 800001 | 910000 |
| WP | X20248_09 | 900001 | 910715 |

Query Match 10.4%; Score 40; DB 1; Length 110000;
Best Local Similarity 46.5%; Pred. No. 1.1;
Matches 155; Conservative 4; Mismatches 165; Indels 9; Gaps 1;

QY 38 AAAATTATCTTAGGGACTGATATTTGTAATTTATGTCATTTAATWRTTKTGGGCAT 97
DB 14020 AAGAGTTTCTAATTGAAAAATTGAGAGATATTATAGTGTTTTATAGATGAGTTTAAGAAT 14079
QY 98 TTCCTTACATTTGCTTGACAAGATTAAATGCTGTGCCAAATTTTGTATTTTAAATTGG 157
DB 14080 AAATTTATTTTGCATTAATAATAGTATATCAAGCATCTTAATGTTTCAGATTTTATCTGA 14139
QY 158 AGACTTCTTATCAAA-----AGTAATGCTGCCAAAGAGTCTAAGGAATTAGTAG 208
DB 14140 TTTATTTATTTTAAACCCAGAGCTAAAAAATTTGCTTAATGATTTGAAGAACAATATAGA 14199
QY 209 TGTCCCMTCACCTTGTGAGAGTGTGCTATTTCTAAAGATTGATTTCCCTGGAATGACA 268
DB 14200 TGTGATGCTCTTGAATTTGAAAGTGTGTTGTTAAAGAAATATGATTCGATTGAAATTTGA 14259
QY 269 ATTATATTTTAACCTTTGCTGGGGGAANAAGTTATAGGACCACAGCTTCACCTTCGATAC 328
DB 14260 AATTAAAAATTTAAATAGACGAGTTGAAAAATCAGATAAAAAATAATTAGTTGAACGC 14319
QY 329 TTGTAATTATCTTTTATTTGACCTGTTTGA 361
DB 14320 AGAATATTTGACTTCTAATTTTGTGTTATATA 14352

RESULT 4

V21209_03/c
Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

Query Match 10.4%; Score 39.8; DB 1; Length 110000;
Best Local Similarity 47.9%; Pred. No. 1.2;
Matches 101; Conservative 4; Mismatches 106; Indels 0; Gaps 0;

QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATAT 60
DB 41592 AAGATCCAGAGACAAAGAACTGTGATTTGAGTAGAAGATACATTTATTTGGAGAGATTG 41533
QY 61 TGGTATATATGTCATTTAATWRTTKTGGGGCATTTCCCTTACATTTGCTTGACAAGA 120
DB 41532 TAGAAATTGAAGCAGATTTAGTTGTAATGTCAGCAGAGATTGTCACCAAGACCAGACAAATC 41473
QY 121 TTAATAATGCTGTGCCAAAAATTTGTAATTTTATTTGGAGACTTCTTATCAAAAGTAATGC 180

Db 41472 CAAATTCGCTAGATGCTTGGTTAGAGCTCAGTCCAGATGATCTTCAAGAGTTGC 41413
OY 181 TGCCAAAGGAAGTCTAAGGAATTAGTAGTGT 211
Db 41412 ATCCAAAGTGTAGCTCCAGTTAATACAAAGGT 41382

RESULT 5
X20248_06
Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 10.3%; Score 39.4; DB 1; Length 110000;
Best Local Similarity 43.5%; Pred. No. 1.5;
Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

OY 29 TCATTCAGAAAAATATCTTAGGAGCTGATATGCTAATTATGCTCAATTTAATWRTRT 88
Db 76486 TGATTCCTTAACCTTATTTGTTAAAGACTTAAAGCTAATGTTAAATTTAAATGTT 76545
OY 89 KTGCGGCATTCCTTACATGTGCTTGACAGATTAAATGCTGTCGCCAAATTTGTAT 148
Db 76546 TCCATCGATATCTCGTCGTAGCTTTCAGATATATTATTGCCCCCAAGTATTTGGAT 76605
OY 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGGAGCTTAAGGAATTAGTAG 208
Db 76606 TTTATTTGTAATGTTATGTAAGAATAATTTTGTAGATAAATCTTGAAGAGCTCTTT 76665
OY 209 TGTCCCMTCACCTTGTGAGTGTGCTATCTTAAAGATTGATTTCTCGAATGACA 268
Db 76666 GAATAAGATAGGTATATGTAATCTCTTTTCTAAAAATCTAATTTAAAAATAAATT 76725
OY 269 ATTATATTTAACTTTGGTGGGGGAANAGTTATAGGACCACAGCTTCACTTCGTATAC 328
Db 76726 TACATGAAAAATATTGTTACTCTTAAACTTAAAGATGTAGATTAGCTCTTTATTTTC 76785
OY 329 TTGTAATTAATCTTTTATTGACACTGTTTGGACCATTAAAGCTAATGTTTAA 382
Db 76786 ATTAAGATCCAACTTTTCTATTGTTTGGCAAACTCTTCTAATTTTAAAGTAA 76839

RESULT 6
X20250
ID X20250 standard; DNA; 111309 BP.
AC X20250;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #3.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,

PI White OR;
DR WPI; 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 738-800; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T;

Query Match 10.3%; Score 39.4; DB 1; Length 111309;
Best Local Similarity 43.5%; Pred. No. 1.5;
Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

OY 29 TCATTCAGAAAAATATCTTAGGAGCTGATATGCTAATTAATGCTCAATTTAATWRTRT 88
Db 87100 TGATTCCTTAACCTTATTTGTTAAAGACTTAAAGCTAATGTTAAATTTAAATGTT 87159
OY 89 KTGCGGCATTCCTTACATGTGCTTGACAGATTAAATGCTGTCGCCAAATTTGTAT 148
Db 87160 TCCATCGATATCTCGTCGTAGCTTTCAGATATATTATTGCCCCCAAGTATTTGGAT 87219
OY 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGGAGCTTAAGGAATTAGTAG 208
Db 87220 TTTATTTGTAATGTTATGTAAGAATAATTTTGTAGATAAATCTTGAAGAGCTCTTT 87279
OY 209 TGTCCCMTCACCTTGTGAGTGTGCTATCTTAAAGATTGATTTCTCGAATGACA 268
Db 87280 GAATAAGATAGGTATATGTAATCTCTTTTCTAAAAATCTAATTTAAAAATAAATT 87339
OY 269 ATTATATTTAACTTTGGTGGGGGAANAGTTATAGGACCACAGCTTCACTTCGTATAC 328
Db 87340 TACATGAAAAATATTGTTACTCTTAAACTTAAAGATGTAGATTAGCTCTTTATTTTC 87399
OY 329 TTGTAATTAATCTTTTATTGACACTGTTTGGACCATTAAAGCTAATGTTTAA 382
Db 87400 ATTAAGATCCAACTTTTCTATTGTTTGGCAAACTCTTCTAATTTTAAAGTAA 87453

RESULT 7
V22738/c
ID V22738 standard; DNA; 2287 BP.
AC V22738;
DT 28-SEP-1998 (first entry)
DE Babesia microti BMNI-8 antigen sequence.
KW antigen; detection; diagnosis; vaccine; tick-borne disease;
KW differentiation; Lyme disease; ehrlichiosis; ss.
OS Babesia microti.
FH key Location/Qualifiers
FT CDS 31..1806 /tag= a
FT /product= antigen
PN EP-834567-A2.
PD 08-APR-1998.
PF 01-OCT-1997; 117067.
PR 24-APR-1997; US-845258.
PR 01-OCT-1996; US-723142.
PA (CORI-) CORIXA CORP.
PI Houghton R, Lodes MJ, Reed SG, Sleath PR;
DR WPI; 98-195465/18.
DR P-PSDB; W56287.
PT Polypeptides comprising Babesia microti antigens and their
PT immunogenic fragments or epitopes - and related nucleic acid,
PT vectors, transformed cells and antibodies, useful for diagnosis of
PT infection and in protective vaccines

PS Claim 8; Page 27-29; 113pp; English.
CC The sequence is that encoding a polypeptide comprising at least
CC one antigenic portion of a Babesia microti antigen. It can be used
CC to diagnose B. microti infection by detecting specific antibodies
CC in usual immunoassays. Infection can also be diagnosed using:
CC (a) primers or probes derived from the coding sequence, in
CC standard amplification or hybridisation tests, or (b) using
CC antibodies to detect the corresponding antigen. It is also
CC useful in vaccines to protect against infection, especially
CC when formulated with an adjuvant. The new diagnostic methods
CC allow rapid differentiation between B. microti infection and
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that
CC have similar symptoms but require different treatments.
SQ Sequence 2287 BP; 981 A; 253 C; 250 G; 803 T;

Query Match 10.1%; Score 38.6; DB 1; Length 2287;
Best Local Similarity 44.1%; Pred. No. 1.5;
Matches 143; Conservative 3; Mismatches 178; Indels 0; Gaps 0;

OY 59 ATGTGTAATATGTCATTTAATWRTTKTGCGGCATTTCCTTACATGCTTGACAA 118
DB 2133 ATTAGTATGATATATGATGACATGAGGTTGATTGTTATACATGTCGATATAT 2074
OY 119 GATTAAGATGCTGCGCAAAATTTGATTTTGGAGACTTCTTATCAAAAGTAAT 178
DB 2073 TATATATATGATATATGATGATGATAGAGATATTTTAAATAGTAT 2014
OY 179 GCTGCCAAGAAGCTAAGAAATAGTAGTGTCCCTCCTGTTGGAGTGTCTAT 238
DB 2013 TAATGATATATATGATATATATATATATGATGATGATGATGATGATGAT 1954
OY 239 TCTAAAGATTTGATTTCTGGAATGACAAATATATTTTAACTTTGGTGGGGAANAG 298
DB 1953 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1894
OY 299 TTATAGGACACAGCTTCTGATCTGTAATTAATCTTTATGTCACCTGTTT 358
DB 1893 TTCTACAAATTAATTTTGTGTAATTAATTAATTAATTAATTAATTAATTA 1834
OY 359 TGACCATTAAGCTAATGTTTAA 382
DB 1833 TAAATATATATATCTCATTA 1810

RESULT 8
V74419

ID V74419 standard; DNA; 4093 BP.

AC V74419;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #108.

KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;
toxic shock syndrome; ds.

KW Staphylococcus aureus.

OS Key Location/Qualifiers

FT misc_feature 421..480

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

FT misc_feature 2221..2280

are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"

FT
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S. aureus vaccines
PS Claim 1; Page 624-627; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S. aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S. aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S. aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 4093 BP; 1261 A; 757 C; 488 G; 1400 T;

Query Match 9.9%; Score 38; DB 1; Length 4093;
Best Local Similarity 51.2%; Pred. No. 2.1;
Matches 86; Conservative 1; Mismatches 81; Indels 0; Gaps 0;

OY 6 AAACAACAAAAAACAATCTCTTCATTCAGAAAAATTAATCTAGGACGATATGTA 65
DB 1633 AAATATAAAAAATAAGATAACACGCTTCATCATCATTTATATGATTAATGTC 1692
OY 66 ATTATGCTCAATTTAATWRTTKTGCGCATTTCTTACATGCTTGCACAAATTA 125
DB 1693 ACTCTCATCAATTTATTTTAAATACACGTTTCATGATTAATTAAGCACGTTCAA 1752
OY 126 ATGTCTGTGCCAAATTTTGTATTTTGGAGACTTCTTATCAAA 173
DB 1753 ATGTAAGTACTGATCTTTATATGTTTAAATGCAATCCATATCAAGA 1800

RESULT 9

ID V01871/c

AC V01871; standard; DNA; 2636 BP.

DT 20-APR-1998 (first entry)

DE Human brain specific nucleosome assembly protein gene.

KW Human; foetal brain cDNA library; GDP dissociation stimulating protein;
brain specific nucleosome assembly protein; diagnosis; therapy;

KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;
ncl-related protein type 1; ncl-related type 2; hereditary disease;

KW Cancer; ss.

KW Homo sapiens.

OS Key Location/Qualifiers

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

FT CDS 266..1786

PA (SAKA) OTSUKA PHARM CO LTD.
PI Fujiwara T, Horie M, Watanabe T;
DR WPI; 97-459830/43.
P-PSDB; W37495.
DR Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT useful for diagnosis or therapy of hereditary disease and cancer
PS Claim 6; Page 60-64; 123pp; English.
CC The present sequence encodes a brain specific nucleosome assembly
CC protein isolated from a human foetal brain cDNA library. The nucleotide
CC or amino acid sequences are useful for in-vitro diagnosis of hereditary
CC diseases and cancer and for preparation of pharmaceuticals.
SQ Sequence 2636 BP; 814 A; 521 C; 599 G; 702 T;

| | | | | |
|---------------------------|--------|-----------------|-----------|--------------|
| Query Match | 9.8% | Score 37.6; | DB 1; | Length 2636; |
| Best Local Similarity | 46.6%; | Pred. No. 2.5; | | |
| Matches 109; Conservative | 3; | Mismatches 122; | Indels 0; | Gaps 0; |

| | | | |
|----|------|---|------|
| OY | 55 | TGATATTGGTAATTATAGTCAATTTAATWRTRTKTGCGGCATTTCCTTACATTTGCTTG | 114 |
| | | : : | |
| Db | 2039 | TGCTTAGATAATTGACAACAGCCTTACATATTTTGAAGAATATTACTACAACAGCCTA | 1980 |
| OY | 115 | ACAAGATTAAATGTCTGTGCCAAAATTTTGTATTTTATTTGGAGACTTCTATCAAAG | 174 |
| | | | |
| Db | 1979 | GATAGAATAAACTGGCACTTAGACACTTTTtagacfaTTTTTAAATATAGACTHCCA | 1920 |
| OY | 175 | TAATGCTGCCAAGAAGACTTAAGGAATTAGTAGTGTCCMTCACCTGTTGGAGTGTG | 234 |
| | | | |
| Db | 1919 | GAAACCACAAAACTGACTATAAGGAATATAGTGTTCAAGTTACAGGTCAGGGTTTTA | 1860 |
| OY | 235 | CTATTCTAAAAGATTTGATTTCCTGGAATGACAATTAATTTAACTTTGGTG | 288 |
| | | | |
| Db | 1859 | CTTTTTCAGGCTGTATGAATCTGCTTCACCTCTTGAGATTTTAAGATTCTTG | 1806 |

RESULT 10
X19427
ID X19427 standard; DNA; 685 BP.
AC X19427;
DT 04-JUN-1999 (first entry)
DE Nicotiana tabacum matrix attachment region PS211-1 DNA sequence.
KW Nicotiana tabacum; matrix attachment region; MAR; tabacco;
KW expression; transgene; structural gene; ss.
OS Nicotiana tabacum.
PN W09907866-A1.
PD 18-FEB-1999.
PF 05-AUG-1998; U16344.
PR 06-AUG-1997; US-066118.
PA (UYNC-) UNIV NORTH CAROLINA STATE.
PI Michalowski SM, Spliker SL;
DR MPI; 99-167436/14.
PT New Matrix Attachment Region DNA - useful for increasing expression
PT of transgene produce, especially structural genes in tobacco
PS Claim 1; Fig 3; 66pp; English.
CC The present sequence represents DNA encoding a Matrix Attachment Region
CC (MAR) (which bind to the nuclear matrix) isolated from Nicotiana tabacum
CC A DNA construct containing MAR sequences can be used to increase
CC expression of a transgene product. A new method identifies MARS by
CC sequence alone, unlike prior art methods using associated genes. MARS
CC increase expression of a transgene product.
SQ Sequence 685 BP; 243 A; 107 C; 87 G; 248 T;

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 9.7%; | Score 37.2; | DB 1; | Length 685; |
| Best Local Similarity | 49.48; | Pred. No. 2.6; | | |
| Matches 87; Conservative | 3; | Mismatches 86; | Indels 0; | Gaps 0; |

OY 31 ATTCAGAAAAATATCTTAGGAGCTGATATGGTAATTATGGTCATTTAATWRTRTTKT 90
 ||| ||||| | ||| ||| ||| ||| ||| : : |||
 Db 366 ATTAATAAATAAGATGGGGCCGCAATTTTATTATTAATAATTATTATTTT 425
 OY. 91 GGGGCATTCCTACATTGTCTTGACACAAGTTAAATGCTGTGCCAAAATTTTGTATTT 150
 | | | | | | | | | | | | | | | | | | | | |

Db 426 AGCGAAGATCCCTCCCTTATTATTGATAACCCTTAACTGACTACATCTTATTATTACT 485

Oy 151 TATTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAAGTCTAAGGAATTAGT 206

Db 486 AAGTTTGCTATAATTATGAAAGTCATCTCTACATACATAAATAATACATATTAAAT 541

RESULT 11
V21209_07
Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

| | | | | |
|-----------------------|-----------------|-----------------|-----------|----------------|
| Query Match | 9.6%; | Score 36.8; | DB 1; | Length 110000; |
| Best Local Similarity | 45.2%; | Pred. No. 5.7; | | |
| Matches 122; | Conservative 3; | Mismatches 145; | Indels 0; | Gaps 0; |

| | | | |
|----|-------|--|--------|
| QY | 18 | AAAACAATCTTCATTCAGAAAAATATCTTAGGAGCTGATATGGTAATTATGCTCAAT | 77 |
| | | | |
| Db | 64033 | ATAAGTATATTTTCACACCTTTATATTTATTTTGCAATTAATTAATTAATTTTAAAT | 640922 |
| QY | 78 | TTAATWRTTTRTKTGGGCGCAATTCCTTACATTCGTCTTGACAGATTAATAATGCTGTGCCA | 137 |
| | | | |
| Db | 64093 | ATTTCTGCACCTTAATATTTTAAATGATTTTTTTGCAAAAAAGAAACATTTTCCGT | 64152 |
| QY | 138 | AAATTTTGATTTTATTTTGAGACCTCTATCAAAAGTAATGCTGCCAAAGGAGCTTAA | 197 |
| | | | |
| Db | 64153 | AAATTTTAAATGAATTTTATAGAAAAATGAAAAATTAATCTTATATAAAAAATAAAAAAT | 64212 |
| QY | 198 | GGAATTAGTAGTGTTCCCMTCACCTGTTTGAGAGTGTGCTATTCATAAAAGATTGATTC | 257 |
| | | | |
| Db | 64213 | TAGAAATTAAGTTAAATTACCAATTTATTTGATTTAAATTTTACCTATCCCTTATTTAT | 642722 |
| QY | 258 | CTGGAATGACAATTATATTTTAACCTTTGGT | 287 |
| | | | |
| Db | 64273 | TCCTTTTGAGACATTAATTTTCACATTTGCT | 64302 |

| RESULT | 12 |
|--------|---|
| ID | Q27886 standard; DNA; 2418 BP. |
| AC | Q27886; |
| DT | 04-FEB-1993 (first entry) |
| DE | P.falciparum GBP130h. |
| KW | Polymerase chain reaction; glycoporin binding protein; inverse PCR, |
| KW | Plasmodium falciparum; malaria; ss. |
| OS | Synthetic. |
| FH | Key |
| FT | exon |
| FT | Location/Qualifiers |
| FT | 767..955 |
| FT | /*tag= a |
| FT | /number= 1 |
| FT | 1111..2202 |
| FT | /*tag= b |
| FT | /number= 2 |
| FT | repeat_region |
| FT | 1249..2202 |


```
FT      EP-499834-A.      /*tag= c
PN      26-AUG-1992.
PD      27-JAN-1992; 101271.
PF      21-FEB-1991; DE-105348.
PR      (BEHW ) BEHRINGWERKE AG.
PI      Hundt E, Knapp B, Kupper H, Nolte D, Ruepper H;
DR      WPI; 92-286009/35.
PT      DNA coding protein GBP 130 h and proteins produced - used for
PS      vaccines to control malaria
CC      Claim 1; Page 12-14; 17pp; German.
CC      Glycophorin binding protein 130h is a P.falciparum blood stage
CC      antigen 69k homologous with the known GBP130. The GBP130h gene and
CC      its homologues can be used in vaccines to protect against malaria
CC      caused by P.falciparum. The coding sequence was isolated using PCR
CC      techniques on a genomic P.falciparum FCBR strain library. The
CC      GBP130h gene was found to be highly conserved between different
CC      strains but is distinct from the GBP130 gene.
CC      See Q27878-Q27885 and Q33432.
SQ      Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T;
```

```
Query Match      9.4%; Score 36; DB 1; Length 2418;
Best Local Similarity 42.6%; Pred. No. 5.6;
Matches 159; Conservative 4; Mismatches 210; Indels 0; Gaps 0;
```

```
QY      1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
DB      638 ATAAAAATAAAAAAAACAATACATAAATATTTTATTTATTTTAAATTTT 579
QY      61 TGGTAATATGTCATTAATWRTTKTGGGGCATTCTTACATTGCTTGACAAGA 120
DB      578 TTTTTTTTCATTAATTAATTCATTCAATGTTTTCATTAAGATAAACTATATAT 519
QY      121 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGGAGACTCTTATCAAAAGTAATGC 180
DB      518 ATATTTTTCATGATATATAATATATCATATCCTTGAAAGAATATATGATATATTATAT 459
QY      181 TGCCAAAGGAAGTCTAAGGAATTAGTAGTGTCCMTCACCTGTTGGAGTGTCTATTC 240
DB      458 TAGTGAAGTTACAAGAAGTGAATAATATATATGATATACACTTACGACTATGATTATA 399
QY      241 TAAAGATTTTGTATTCCTGGAATGACAATTATATTAACTTTGGTGGGGAANAAGTT 300
DB      398 AATAATATTAATATTCATGATATATATTTTATTTGAATCTTAAATATTAATAAGAA 339
QY      301 ATAGGACACAGTCTTCACCTCTGACTCTGTAATTAATCTTTTATGCACTGTTTG 360
DB      338 TTAGAATATATTTTATATATATATTTTCCACATGTTCTTTGTATCTATCTTTTTC 279
QY      361 ACCATTAAAGCTAT 373
DB      278 TATTATAAATAT 266
```

RESULT 13

```
T89990/c
ID      T89990 standard; DNA; 3607 BP.
AC      T89990;
DT      11-MAY-1998 (first entry)
DE      Cryptosporidium parvum oocyst heat shock protein HSP70 gene.
KW      Cryptosporidium; detection; assay; protozoan; heat shock protein;
OS      HSP70; gene amplification; PCR; ds.
FH      Cryptosporidium parvum.
FT      Key Location/Qualifiers
FT      CDS 1272..3296
FT      /*tag= a
FT      /transl_except= (pos:1536..1538, aa:His)
FT      /transl_except= (pos:1578..1580, aa:Leu)
FT      /transl_except= (pos:1659..1661, aa:Lys)
FT      /transl_except= (pos:2034..2036, aa:Lys)
FT      /transl_except= (pos:2265..2267, aa:Leu)
FT      /transl_except= (pos:2931..2933, aa:Lys)
```

```
FT      /transl_except= (pos:3093..3095, aa:Lys)
FT      /transl_except= (pos:3105..3107, aa:His)
FT      /transl_except= (pos:3267..3269, aa:Lys)
PN      WO9742349-A1.
PD      13-NOV-1997.
PF      08-MAY-1997; U07972.
PR      09-MAY-1996; US-647351.
PA      (METR-) METROPOLITAN WATER DISTRICT SOUTHERN CALIFORNIA.
PI      De Leon R, Rochelle PA;
DR      WPI; 97-55899/51.
DR      P-PSDB; W31345.
```

```
PT      Amplifying and detecting viable and infective protozoan oocysts -
PT      by monitoring presence or absence of heat shock protein genes in
PT      their transcripts
PS      Claim 12; Page 23-26; 44pp; English.
CC      This genomic DNA sequence comprises the heat shock protein HSP70
CC      gene of Cryptosporidium parvum oocysts. The invention provides
CC      processes and kits for detecting encysted forms of protozoa,
CC      particularly Cryptosporidium (especially Cryptosporidium muris,
CC      Cryptosporidium baileyi, Cryptosporidium wrairi and C. parvum)
CC      and Giardia, that are viable and infectious by the enzymatic
CC      amplification (e.g. by PCR) of a target gene sequence such as HSP70.
CC      A preferred primer pair (see T89991-92) can amplify DNA or RNA
CC      from many different Cryptosporidium species, and can be used with a
CC      probe (see T89997) that confirms the identity of the amplification
CC      product. A most preferred primer pair (see T89993-94) is specific
CC      to C. parvum and can also be used with a probe (see T89998) to
CC      confirm sequence identity. In addition, both Cryptosporidium and
CC      Giardia can be detected simultaneously using a multiplex
CC      amplification reaction. The methods exploit the speed, sensitivity
CC      and specificity associated with an amplification procedure, and
CC      enable pathogenic forms of protozoa present in low copy numbers to
CC      be identified and distinguished from morphologically similar, but
CC      non-pathogenic protozoa.
SQ      Sequence 3607 BP; 1265 A; 514 C; 716 G; 1112 T;
```

```
Query Match      9.4%; Score 36; DB 1; Length 3607;
Best Local Similarity 48.3%; Pred. No. 5.8;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;
```

```
QY      1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTAATCTTAGGACTGATAT 60
DB      221 AATAAAAAAAATATACAACTTTTATTTAATAAATAATCTTAATAAGTGAATA 162
QY      61 TGGTAATATGTCATTAATWRTTKTGGGGCATTCTTACATTGCTTGACAAGA 120
DB      161 TGGATCTCCTCATCAATCAATAATATTTTACAGACTAATCAACAATATGATTATTA 102
QY      121 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGGAGACTTCTTAACAAAGTAATGC 180
DB      101 TGAACATAATGCTGTGAATAATTAATTTCAATAGGAAAAAATATTTTATTAATATTTAC 42
```

RESULT 14

```
V21209_05
Continuation (6 of 17) of V21209 from base 500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP      Fragment Name      Begin      End
WP      V21209_00          1          110000
WP      V21209_01          100001     210000
WP      V21209_02          200001     310000
WP      V21209_03          300001     410000
WP      V21209_04          400001     510000
WP      V21209_05          500001     610000
WP      V21209_06          600001     710000
WP      V21209_07          700001     810000
WP      V21209_08          800001     910000
WP      V21209_09          900001    1010000
WP      V21209_10         1000001    1110000
WP      V21209_11         1100001    1210000
WP      V21209_12         1200001    1310000
WP      V21209_13         1300001    1410000
```


| | | | |
|----|-----------|---------|---------|
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

| | | | | |
|---------------------------|--------|-----------------|-----------|----------------|
| Query Match | 9.38; | Score 35.8; | DB 1; | Length 110000; |
| Best Local Similarity | 43.0%; | Pred. No. 9.5; | | |
| Matches 154; Conservative | 3; | Mismatches 201; | Indels 0; | Gaps 0; |

| | | | |
|----|-------|---|-------|
| OY | 25 | TTCCTCATTCAGAAAAATTATCTTAGGGACGATAITGGTAATATATGGTCAATTAAATMR | 84 |
| Db | 96810 | TACTTCATAATAAATAAGTGCCCTTTTTCCTAATGTGTGAATTAATAGTATGTTTTGC | 96869 |
| OY | 85 | TRTKTGCGGCATTTCTCCTTACATTTGCTTGACAAGATTAAAATGTCTGTGCCAAAATTTT | 144 |
| Db | 96870 | TAAIGCCTGTTGTGTCTCTTTCTTTTGATTAATTAATCCAATTTTACCNAATTCATT | 96929 |
| OY | 145 | GTAATTTATTTGGAGACTTCTTATCAAAGAATAATGCTGCCAAGAGACTTAAGAAATTA | 204 |
| Db | 96930 | TAACTCTGTTTAAACTGCTCATATAGTAGGAATTTCTTATGAATTTGACTAAAATG | 96989 |
| OY | 205 | GTAGTGTCCCMTCACTTGTGTGGAGTGTGCTATTTCTAAAAGATTTTGATTTTCCGTGAAT | 264 |
| Db | 96990 | GAGTTCTCTCTTTTCATTAAAGTAGTGTCCAAAATTTCTTAAACATAATTTTTCCTAGTAT | 97049 |
| OY | 265 | GACATTTATATTTAACCTTGGTGGGGGAANAAGTTATAGGACCACAGCTTCACTTCTG | 324 |
| Db | 97050 | CACCAATAAGCATGATTTAGCCCCCTTTATGAATTTTATTTTTCGCTTTTGGAGAATTTTC | 97109 |
| OY | 325 | ATACTGTAAATTAATCTTTTATTTGACACTGTTTTGACCATTAAAGCTAATATGTTAA | 382 |
| Db | 97110 | AATTTATACGATAATTAACATTTTTCGCTAAATTTTCCAAAAAAGAGAAAAGTTATA | 97167 |

RESULT 15

X12949/c
ID X12949 standard; DNA; 9047 BP.

DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig ID NO:12.
KM Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.

PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI BARASH SC, DILLON PJ, KUNSCH CA.;
DR WPI; 99-045171/04.

PT New isolated *Enterococcus faecalis* polynucleotides and polypeptides
PT - used to develop products for the detection of *Enterococcus* and for
PT use in vaccines for prevention or attenuation of *Enterococcus*
PT infection.

PS Claim 1; Page 305-310; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it
CC 882 nucleotide sequences isolated from the *Enterococcus faecalis* genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the *Enterococcus faecalis* genome with
CC commercial importance. The products can be used to detect the presence
CC of *Enterococcus faecalis* in samples. They can also be used for
CC diagnosing *Enterococcal* infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or
CC another related organism, *in vivo* or *in vitro*. In particular the
CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
CC can be used in vaccines to prevent or attenuate an *Enterococcal*
CC infection.
SO Sequence 9047 BP; 3107 A; 1542 C; 1841 G; 2551 T;

| | | | | |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match | 9.38; | Score 35.6; | DB 1; | Length 9047; |
| Best Local Similarity | 50.68; | Pred. No. 7.9; | | |
| Matches 80; | Conservative 2; | Mismatches 76; | Indels 0; | Gaps 0; |

```
QY      53 ACIGATATTGGTAATTATGCTCAATTTAATWRTKTTGGGCATTTCCTTACATTGTCT   112
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     6436 AGTCAACTTATTAAATGCGTTTCTATGAATAAATCAAGGTGCCATTACTTTTGATGGAT   6377
QY      113 TGACAAGATTAAAATGTCGTGCCAAAAATTTTGTATTTATTTGAGACTTCTATCAA   172
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     6376 TGATATTACGAACCTTCTCGACAAAAATCTAAGAATTTATTTGGCATGSTTTTACAAA   6317
QY      173 AGTAATGCTGCCAACGAAGAGTCTAAGGAATTAGTAGTG   210
        ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db     6316 CACTTGCGTATTGGAAGGAACCGTAGCAGATAATATTG   6279
```

Search completed: September 28, 1999, 12:30:02
Job time: 5422 sec

Job time: 5422 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:34:11 ; Search time 161.06 Seconds
(without alignments)
218.006 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383

Sequence: 1 AAACAACAACAACAAAAA.....ATTACCTATATGTTTAA 383

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/1na/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/2/1na/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/2/1na/PCTUS9_COMB.seq:*
- 6: /cgn2_6/ptodata/2/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 37.6 | 9.8 | 2636 | 4 | US-08-820-170A-21 |
| 2 | 36 | 9.4 | 3607 | 3 | US-08-647-351B-1 |
| 3 | 34.8 | 9.1 | 5852 | 1 | US-07-867-106-2 |
| 4 | 34.6 | 9.0 | 2334 | 2 | US-08-062-632-4 |
| 5 | 34.2 | 8.9 | 2749 | 1 | US-08-344-695-1 |
| 6 | 34 | 8.9 | 5852 | 1 | US-07-867-106-2 |
| 7 | 33.6 | 8.8 | 8920 | 4 | US-08-446-855A-1 |
| 8 | 31.6 | 8.3 | 2074 | 2 | US-08-414-685-1 |
| 9 | 31.6 | 8.3 | 1422 | 3 | US-08-319-704-5 |
| 10 | 31.4 | 8.2 | 2636 | 1 | US-08-253-785-1 |
| 11 | 31.2 | 8.1 | 901 | 3 | US-08-832-883-65 |
| 12 | 31.2 | 8.1 | 901 | 4 | US-08-832-883-65 |
| 13 | 30.8 | 8.0 | 7101 | 3 | US-08-480-604A-9 |
| 14 | 30.6 | 8.0 | 414 | 1 | US-08-377-687-48 |
| 15 | 30.6 | 8.0 | 8878 | 1 | US-08-206-176-3 |
| 16 | 30.6 | 8.0 | 414 | 3 | US-08-777-192-48 |
| 17 | 30.6 | 8.0 | 1779 | 5 | PCT-US96-05320A-831 |
| 18 | 30.4 | 7.9 | 1939 | 1 | US-07-715-751B-2 |
| 19 | 30.4 | 7.9 | 1511 | 1 | US-07-991-867B-8 |
| 20 | 30.4 | 7.9 | 660 | 1 | US-07-991-867B-32 |
| 21 | 30.4 | 7.9 | 3661 | 1 | US-08-105-483-221 |
| 22 | 30.4 | 7.9 | 3661 | 2 | US-08-303-124-7 |
| 23 | 30.4 | 7.9 | 3661 | 2 | US-08-204-729-7 |
| 24 | 30.4 | 7.9 | 2762 | 2 | US-08-198-446B-12 |
| 25 | 30.4 | 7.9 | 1511 | 2 | US-08-107-755A-8 |
| 26 | 30.4 | 7.9 | 660 | 2 | US-08-107-755A-32 |
| 27 | 30.4 | 7.9 | 3660 | 3 | US-08-475-063-25 |
| 28 | 30.4 | 7.9 | 3660 | 3 | US-08-207-792-25 |
| 29 | 30.4 | 7.9 | 3661 | 3 | US-08-709-209-221 |
| 30 | 30.4 | 7.9 | 3661 | 3 | US-08-458-101-221 |
| 31 | 30.4 | 7.9 | 2762 | 4 | US-08-870-693-12 |
| 32 | 30.4 | 7.9 | 4951 | 5 | PCT-US95-06119-5 |
| 33 | 30.2 | 7.9 | 1560 | 3 | US-08-356-180-1 |
| 34 | 30.2 | 7.9 | 2213 | 3 | US-08-592-936B-12 |
| 35 | 30.2 | 7.9 | 2145 | 3 | US-08-592-936B-16 |
| 36 | 30.2 | 7.9 | 2917 | 3 | US-08-592-936B-20 |
| 37 | 30.2 | 7.9 | 1356 | 3 | US-08-592-936B-22 |

| | | | | | | |
|----|------|-----|--------|---|-------------------|-------------------|
| 38 | 30.2 | 7.9 | 2213 | 3 | US-08-788-928A-1 | Sequence 1, Appl |
| 39 | 30.2 | 7.9 | 1356 | 3 | US-08-788-928A-2 | Sequence 2, Appl |
| 40 | 30.2 | 7.9 | 259 | 3 | US-08-788-928A-4 | Sequence 4, Appl |
| 41 | 30.2 | 7.9 | 246240 | 4 | US-08-724-394A-20 | Sequence 20, Appl |
| 42 | 30.2 | 7.9 | 246240 | 4 | US-08-724-394A-21 | Sequence 21, Appl |
| 43 | 30.2 | 7.9 | 246240 | 4 | US-08-724-394A-22 | Sequence 22, Appl |
| 44 | 30 | 7.8 | 731 | 3 | US-08-451-405A-2 | Sequence 2, Appl |
| 45 | 30 | 7.8 | 246240 | 4 | US-08-724-394A-22 | Sequence 22, Appl |

ALIGNMENTS

RESULT 1
US-08-820-170A-21/c
Sequence 21, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
LIBRARY: Human fetal brain cDNA library
CLONE: GEN-078D05
FEATURE:
NAME/KEY: CDS
LOCATION: 266..1783
US-08-820-170A-21
Query Match 9.8%; Score 37.6; DB 4; Length 2636;
Best Local Similarity 46.6%; Pred. NO. 0.27;
Matches 109; Conservative 3; Mismatches 122; Indels 0; Gaps 0;
OY 55 TGATATGGTAATATGTCATTTAATWRTTKTGGGCATTTCCCTACATGCTTG 114
DB 2039 TGCTTAGATAATGACACACGCTACATATTTGAAGATATATACACAGCCCTA 1980
OY 115 ACAGATTAAATGTCTGCGCCAAATTTGTAATTTTGGAGACTTCTATCAAAAG 174
DB 1979 GATGAGATAAAGTGGCCTTAGACACTTTTGAAGACTATTTTAAATATAGACTACCA 1920

APPLICANT: Minion, F. Chris
TITLE OF INVENTION: PCR-Based Assay For Mycoplasma
TITLE OF INVENTION: Hypopneumoniae
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dickstein, Shapiro and Morin
STREET: 2101 L. St. NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 22037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/062,632
FILING DATE: 18-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brady Jr., James W.
REGISTRATION NUMBER: 32,115
REFERENCE/DOCKET NUMBER: 18900.018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)785-9700
TELEFAX: (202)887-0689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2334 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-632-4

Query Match 9.0%; Score 34.6; DB 2; Length 2334;
Best Local Similarity 48.6%; Pred. No. 1.4;
Matches 85; Conservative 3; Mismatches 87; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
DB 496 AAAATAAGGAAGAAGATATTTAAACATGGAATAATATGGAATATTGCAATAT 437
OY 61 TGGTAATATGTCATTTAATRTTKTGGGCATTCCTTACATTCCTTGACAAGA 120
DB 436 CAGAAATAAGGTCACCTCTTACTAAATCAATATTTACTATAGAAATTTATATTA 377
OY 121 TTAATAATGCTGTGCCAAATTTGTATTTATTGGAGACTTCTTATCAAAAGT 175
DB 376 TTAGTATGTTTGCCCTCAAAATTCAGCACTTTTATATATAAATCAAAATTAAT 322

RESULT 5
US-08-344-695-1/c
Sequence 1, Application US/08344695
Patent No. 5614398
GENERAL INFORMATION:
APPLICANT: O'BROCHTA, DAVID
APPLICANT: WARREN, WILLIAM
APPLICANT: ATKINSON, PETER
TITLE OF INVENTION: A GENE TRANSFER SYSTEM FOR INSECTS
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/344,695
FILING DATE: 18-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kelber, Steven B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-058-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 24855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2749 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 450..2285
US-08-344-695-1

Query Match 8.9%; Score 34.2; DB 1; Length 2749;
Best Local Similarity 49.2%; Pred. No. 1.9;
Matches 87; Conservative 1; Mismatches 89; Indels 0; Gaps 0;

OY 14 AAAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATATGTAATATGCT 73
DB 1147 AAAAAATGCTTTAAGCTCTTATAAATATTTCTGCTGTGAGATCTTCAAAATCTAAG 1088
OY 74 CAATTAATWRTTKTGGGCAATTCCTTACATTCCTTGACAAGAATTAATGCTGT 133
DB 1087 GACTTTAAACCTAAATATAGATCTGCAGTTCATGTTTTCATGTAAGTACT 1028
OY 134 GCCAAATTTGTATTTATTTGAGAGACTTCTTATCAAAAGTATGCTGCCAAGA 190
DB 1027 CCCAAAAATTCGCTTATATATATATATGCGGTCACCAATCTATCGTCTGATGCA 971

RESULT 6
US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187
 APPLICATION NUMBER: PCT/AU90/005300
 FILING DATE: 02-NOV-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Feeney, Joanne Longo
 REGISTRATION NUMBER: 35,134
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5852 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2378..5038
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2378..5038

| | | | | |
|-----------------------|--------|----------------|--------|----------------|
| Query Match | 8.9%; | Score 34; | DB 1; | Length 5852; |
| Best Local Similarity | 48.8%; | Pred. No. 2.4; | | |
| Matches | 82; | Conservative | 3; | Mismatches 83; |
| | | | Indels | 0; |
| | | | Gaps | 0 |

```
QY      1 AAACAACAACAAAAAACATCTCTCATTCAGAAAATTATCTTAGGCACTGATAT 60
        ||||| ||| ||| ||||| ||| | | | | | | | | | |
Db      5626 AAAAAAAAAAAAAAAAAAAAAAAAAATAAAATGTATAAATTGGCAATTAAAAATAAAA 5567
QY      61 TGGTAATTATGCTCAATTTAATWTRRTKTTGGGGCAATTCCTTACATTTGCTTGACAAGA 120
        | | | | | : | : | : | : | | | | | | | | | |
Db      5566 AATGTAAGCGTTTTTTTTTAAATATGATCAGATTTTATTTTAAAAATCATTTGACGAGA 5507
QY      121 TTAAATATGCTGTGCCAAAATTTTGATTTTATTTGGAGACTTCTTAT 168
        ||||| ||| ||| ||||| ||||| ||||| ||| ||| |
Db      5506 TTAATAAATCTTTAAACAATTAACAATATTGATTTTTTTTTTTTTTTT 5459
```

RESULT 7
US-08-446-

; Sequence 1, Application US/08446855A
; Patent No. 5849573

```

; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor

; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA

```

; ZIP: 22201-4714
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 05-Jul-2008

```

CLASSIFICATION: 4355
FILING DATE: 06-JUL-74

ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C

```

;
;   REGISTRATION NUMBER: 29.009
;   REFERENCE/DOCKET NUMBER: 47-80
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-816-4000
;   TELEFAX: 703-816-4100
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 8920 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;   MOLECULE TYPE: genomic
;
US-08-446-855A-1

```

| | | | | |
|-----------------------|--------|----------------|-------|----------------|
| Query Match | 8.88; | Score 33.6; | DB 4; | Length 8920; |
| Best Local Similarity | 59.3%; | Pred. No. 3.3; | | |
| Matches | 51; | Conservative | 3; | Mismatches 32; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|-----|--|-----|
| OY | 1 | AAACCAACAACAAAAAACAATCTTCATTCCAGAAAATATCTTAGGACTGATAT | 60 |
| | | | |
| Db | 633 | AAAAAAAAAAAAAGAAAAAATAACATATGA AAAATGCACTTGATATGTA AAT | 692 |
| OY | 61 | TGGTAAATTATGCTCAATTTAATWRTR | 86 |
| | | . : : : : | |
| Dd | 693 | TTATAAATATTTTAAACATAAATATA | 718 |

RESULT

; Sequence 1, Application US/08414685
; Patent No. 5667989

```

; GENERAL INFORMATION:
; APPLICANT: McCallough, John
; APPLICANT: Baymiller, Judy
; TITLE OF INVENTION: Fungal Cell Wall Protein CLY4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000

CITY: Princeton
STATE: New Jersey
COUNTRY: USA

ZIP: 08543-4000
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: DOS

```

```

;
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;
;   CURRENT ADDITION:  DATA

```

```

CORRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/414,685
:
: FILING DATE:

```

FILED DATE: 530
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33.111

REFERENCE/DOCKET NUMBER: DC333
TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526

```

; INFORMATION FOR SEQ ID NO: 1
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 2074 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic,
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 354..1193

```

US-08-414-685-1

RESULT 11

US-08-832-883-65/c
Sequence 65, Application US/08832883
Patent No. 5807681

GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
APPLICANT: Baldi, Alphonso
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
OF CANCER
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,883
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-8383

TELEFAX: (215) 568-5549

INFORMATION FOR SEQ ID NO: 65:

SEQUENCE CHARACTERISTICS:

LENGTH: 901 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-832-883-65

Query Match 8.1%; Score 31.2; DB 3; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2;
Matches 114; Conservative 4; Mismatches 122; Indels 2; Gaps 1;

QY 20 AACAACTCTCATTCAGAAAAATATCTTAGGAGCTGATATGGTAATTAGTCAATT 79
DB 774 AATACTTAACCTTAAGAAAAATATCTTAATAGCAAACTTAAGTCTTAGTTGTGCCA 715
QY 80 AATWRTTKTGGGGCATTCCTTACAT--TGCTTGACAAGATTAAATGCTGTGCCA 137
DB 714 GTATATATGTGAACAATTTTATACATATGTCATGTCCTTACACACCAACCATAT 655
QY 138 AAATTTGATTTTATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGTCTAA 197
DB 654 GCTAGAACTAGTATATATCCATTTTATATATGAGAAATGAGGTACAGAGAAAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCMTCACCTTGTGAGTGTGCTATTCTAAAGATTGTATTC 257
DB 594 ATGATTTGACAGGGTTACCACTGTTAAGTACTGGGAAATTTGAACCCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 12

US-08-832-877-65/c
Sequence 65, Application US/08832877
Patent No. 5840506

GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,877
FILING DATE:

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-8383

TELEFAX: (215) 568-5549

INFORMATION FOR SEQ ID NO: 65:

SEQUENCE CHARACTERISTICS:

LENGTH: 901 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-832-877-65

Query Match 8.1%; Score 31.2; DB 4; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2;
Matches 114; Conservative 4; Mismatches 122; Indels 2; Gaps 1;

QY 20 AACAACTCTCATTCAGAAAAATATCTTAGGAGCTGATATGGTAATTATGTCATTT 79
DB 774 AATACTTAACCTTAAGAAAAATATCTTAATAGCAAACTTAAGTCTTAGTTGTGCCA 715
QY 80 AATWRTTKTGGGGCATTCCTTACAT--TGCTTGACAAGATTAAATGCTGTGCCA 137
DB 714 GTATATATGTGAACAATTTTATACATATGTCATGTCCTTACACCAACCATAT 655
QY 138 AAATTTGATTTTATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGTCTAA 197
DB 654 GCTAGAACTAGTATATCCATTTTATATATGAGAAATGAGGTACAGAGAAAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCMTCACCTTGTGAGTGTGCTATTCTAAAGATTGTATTC 257
DB 594 ATGATTTGACAGGGTTACCACTGTTAAGTACTGGGAAATTTGAACCCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 13

US-08-480-604A-9
Sequence 9, Application US/08480604A
Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.

;; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
;; NUMBER OF SEQUENCES: 32
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MEDLEN & CARROLL, LLP
;; STREET: 220 MONTGOMERY STREET, SUITE 2200
;; CITY: SAN FRANCISCO
;; STATE: CALIFORNIA
;; COUNTRY: UNITED STATES OF AMERICA
;; ZIP: 94104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/480,604A
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/422,711
;; FILING DATE: 14-APR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/405,496
;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/329,154
;; FILING DATE: 25-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/161,907
;; FILING DATE: 02-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/985,321
;; FILING DATE: 04-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/429,791
;; FILING DATE: 31-OCT-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: INGOLIA, DIANE E.
;; REGISTRATION NUMBER: 40,027
;; REFERENCE/DOCKET NUMBER: OPHD-01763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 705-8410
;; TELEFAX: (415) 397-8338
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7101 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..7098
;; US-08-480-604A-9

Query Match 8.0%; Score 30.8; DB 3; Length 7101;
Best Local Similarity 51.8%; Pred. No. 15;
Matches 59; Conservative 4; Mismatches 51; Indels 0; Gaps 0;

QY 8 ACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACATGATTTGCTAAT 67
DB 4227 AGAAGGAATTAATGCAATATAGAGTGTGATTTATCTTAATCATATAATTACTTAT 4286

QY 68 TATGTCATTTATWTRTTRTKTGGGGCATTTCTTACATTCCTTGACACAAGAT 121
DB 4287 TTCTGGCAATTAATAATATGATGTTAAATTCAAATCATATTCACAGAAAAAT 4340

RESULT 14
US-08-377-687-48/C
; Sequence 48, Application US/08377687

;; Patent No. 5538525
;; GENERAL INFORMATION:
;; APPLICANT: BROEKAERT, WILLEM F.
;; APPLICANT: CAMUE, BRUNO P.A.
;; APPLICANT: OSBORN, RUPERT W.
;; APPLICANT: REES, SARAH B.
;; APPLICANT: TERRAS, FRANKY R.G.
;; APPLICANT: VANDERLEYDEN, JOZEF
;; TITLE OF INVENTION: BIOCIDAL PROTEINS
;; NUMBER OF SEQUENCES: 59
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN DABY & CUSHMAN
;; STREET: 1100 NEW YORK AVENUE, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/377,687
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/002,480
;; FILING DATE: 04-JAN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; INFORMATION FOR SEQ ID NO: 48:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 414 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 16..255
;; US-08-377-687-48

Query Match 8.0%; Score 30.6; DB 1; Length 414;
Best Local Similarity 46.9%; Pred. No. 9.8;
Matches 84; Conservative 3; Mismatches 92; Indels 0; Gaps 0;

QY 20 AACAAATCTTCATTCAGAAAAATATCTTAGGACATGATTTGTAATTAATGCTCAATT 79
DB 391 AATAAACTTATTTGTATTAACCGAACACATTAACATATCTGTACATGCTTAATAATC 332

QY 80 AATWTRTTRTKGGGCATTTCTTACATTCCTTGACACAAGATTAATGCTGTGCCAAA 139
DB 331 ACTCATGATAGTAGACACTGACTTATTTGTGTAATTAATTAATAAATATTCACCAAG 272

QY 140 ATTTGATTTTATTTTGAGACTCTTATCAAAAGTAATGCTGCCAAAGGAGTCTAAG 198
DB 271 AGTTGGCAATTAATTAACAAGAAAGTAGACAGATACACTTGTGAGCTGGGAGACATAG 213

RESULT 15
US-08-206-176-3
; Sequence 3, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A

APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of Fibrinogen in Transgenic
TITLE OF INVENTION: Animals
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-15
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8878 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: human fibrinogen B-beta chain
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..469
FEATURE:
NAME/KEY: exon
LOCATION: 470..583
FEATURE:
NAME/KEY: intron
LOCATION: 584..3257
FEATURE:
NAME/KEY: exon
LOCATION: 3258..3449
FEATURE:
NAME/KEY: intron
LOCATION: 3450..3938
FEATURE:
NAME/KEY: exon
LOCATION: 3939..4122
FEATURE:
NAME/KEY: intron
LOCATION: 4123..5042
FEATURE:
NAME/KEY: exon
LOCATION: 5043..5270
FEATURE:
NAME/KEY: intron
LOCATION: 5271..5830
FEATURE:
NAME/KEY: exon
LOCATION: 5831..5944
FEATURE:
NAME/KEY: intron
LOCATION: 5945..6632
FEATURE:
NAME/KEY: exon
LOCATION: 6633..6758

FEATURE:
NAME/KEY: intron
LOCATION: 6759..6966
FEATURE:
NAME/KEY: exon
LOCATION: 6967..7252
FEATURE:
NAME/KEY: intron
LOCATION: 7253..7870
FEATURE:
NAME/KEY: exon
LOCATION: 7871..8102
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 8103..8537
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 8538..8878
FEATURE:
NAME/KEY: CDS
LOCATION: join(470..583, 3258..3449, 3939..4122, 5043..5270,
5831..5944, 6633..6758, 6967..7252, 7871..8102)
US-08-206-176-3

Query Match 8.0%; Score 30.6; DB 1; Length 8878;
Best Local Similarity 52.3%; Pred. No. 18;
Matches 57; Conservative 4; Mismatches 48; Indels 0; Gaps 0;
QY 42 TTATCTTAGGACTGATATTTGTAATTAATGTCATTTAATWRRRTTKTGGGCATTTCC 101
Db 3553 TTATTTTGTGTTGTTTATTTTGAAATAAAATTCAAAACATAAACAATTTGGCCTTGGT 3612
QY 102 TTACATGTCCTTGACAAAGATTAAATGCTGTGCCAAATTTTGTAATT 150
Db 3613 TTAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTGGTCTGGGCCCAAAATTCAAATT 3661

Search completed: September 28, 1999, 11:34:15
Job time: 2135 sec

This Page Blank (usptc)

DT 28-OCT-1991 (first entry)
DE PreTGF-beta gene product.
KW Transforming growth factor beta; cancer; wound healing.
FH Key Location/Qualifiers
FT protein 279..390
PD EP-200341-A.
PN 10-DEC-1986.
PF 21-MAR-1986; 302112.
PR 22-MAR-1985; US-715142.
PR 13-MAR-1987; US-025423.
PA (GETH) GENENTECH INC.
PI Derynck RMA;
DR WPI; 86-326875/50.
DR N-PSDB; N60972.
PT TGF-beta prodn. from transformed hosts - useful esp. for treating
PS wounds (J6 2/9/86).
CC Disclosure; Fig 1b; 26pp; English.
CC The gene product is known to stimulate cell proliferation and
CC inhibit anchorage-dependent growth of a variety of human cancer cell
CC lines, it is esp. useful in treatment of burns and the promotion of
CC surface and internal wound healing. TGF-beta may be expressed from a
CC transformed CHO cell line.
SQ Sequence 390 AA;

Query Match 20.4%; Score 90; DB 3; Length 390;
Best Local Similarity 43.8%; Pred. No. 1.30e+01;
Matches 14; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Db 6 lrlplllplllvltpppaaglstcktd 37
11 lllslhlglgwldlqepapadelssgkmae 42
QY

RESULT 10
ID R05258 standard; protein; 390 AA.
AC R05258;
DT 05-AUG-1990 (first entry)
DE Human pre-transforming growth factor-beta-1 (pre-TGF-beta-1)
KW Transforming growth factor-beta-1 (TGF-beta-1);
KW neoplastic cell line inhibition;
KW EGF-potentiated anchorage-independent growth.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..278
FT protein 279..2011
FT domain 8..23
FT modified_site /note="hydrophobic domain"
FT modified_site 82..84
FT modified_site /note="potential N-glycosylation site"
FT modified_site 136..138
FT modified_site /note="as above"
FT modified_site 176..178
FT modified_site /note="as above"
FT cleavage_site 277..278
FT /note="proteolytic cleavage site"
PN US4886747-A.
PD 12-DEC-1989.
PF 13-MAR-1987; U25423.
PR 13-MAR-1987; US-025423, US-715142.
PA (GETH) Genentech Inc.
PI Derynck RMA, Goeddel DV;
DR WPI; 90-051338/07.
DR N-PSDB; Q93301.
PT Nucleic acid encoding transforming growth factor-beta -
PT cloned into expression vectors for expression in eukaryotic host
PT cells for therapeutic use
PS Disclosure; Fig 1b; 26pp; English.
CC Pre-TGF-beta-1 is a polypeptide of 390 amino acids. Post-translational
CC cleavage of the precursor gives rise to the mature TGF-beta monomer.
CC The sequence for human TGF-beta was determined by direct amino acid
CC sequence analysis and by deduction from the TGF-beta cDNA. It is
CC capable of inducing EGF-potentiated anchorage-independent growth of
CC target cell lines, and/or growth inhibition of neoplastic cell lines. It

CC can be used for treating wounds, eg burns or epidermal ulcers.
SQ Sequence 390 AA;

Query Match 20.4%; Score 90; DB 1; Length 390;
Best Local Similarity 43.8%; Pred. No. 1.30e+01;
Matches 14; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Db 6 lrlplllplllvltpppaaglstcktd 37
11 lllslhlglgwldlqepapadelssgkmae 42
QY

RESULT 11
ID R46227 standard; protein; 390 AA.
AC R46227;
DT 09-JUL-1994 (first entry)
DE Human pre-TGF-beta-1.
KW TGF-beta-1; TGF-beta-2; transforming growth factor beta-1;
KW transforming growth factor beta-3; recombinant; wound healing;
KW vulnerable.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 279..390
FT cleavage_site /label= Mat_peptide
FT modified_site 279
FT modified_site /note= "TGF-beta-1 release site"
FT modified_site 82..84
FT modified_site /label= N-glycosylation_site
FT modified_site 136..138
FT modified_site /label= N-glycosylation_site
FT modified_site 176..178
FT modified_site /label= N-glycosylation_site
PN US5284763-A.
PD 08-FEB-1994.
PF 22-MAR-1985; 715142.
PR 22-MAR-1985; US-715142.
PR 13-MAR-1987; US-025423.
PR 04-AUG-1989; US-389929.
PR 04-MAR-1992; US-845893.
PA (GETH) GENENTECH INC.
PI Derynck RMA, Goeddel DV;
DR WPI; 94-056343/07.
DR P-PSDB; Q56923.
PT Nucleic acid sequences encoding transforming growth factor-beta -
PT diagnostic probes, and for use in therapeutics
PS Disclosure; Fig 1b; 25pp; English.
CC cDNA sequences were determined for human pre-TGF-beta-1 (Q56923),
CC pig TGF-beta-3 (Q56925) and human TGF-beta-3 (Q56926), and the
CC corresponding amino acid sequences were determined (R46227-29,
CC respectively). A genomic fragment corresponding to a human TGF-
CC beta-1 exon (Q56924) was also isolated and its amino acid sequence
CC determined (R46230). The sequences have been used in the
CC construction of vectors for the expression of recombinant TGF-
CC beta.
SQ Sequence 390 AA;

Query Match 20.4%; Score 90; DB 9; Length 390;
Best Local Similarity 43.8%; Pred. No. 1.30e+01;
Matches 14; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Db 6 lrlplllplllvltpppaaglstcktd 37
11 lllslhlglgwldlqepapadelssgkmae 42
QY

RESULT 12
ID R73596 standard; protein; 390 AA.
AC R73596;
DT 20-DEC-1995 (first entry)
DE Human TGF-beta 1 protein.
KW Transforming growth factor-beta; Human TGF-beta protein; TGF-beta 1;
KW TGF-beta 3; osteogenic cell source; OCS; bone deficiency;
KW bone-inducing cofactor.

OS Homo sapiens.
PN US5409896-A.
PD 25-APR-1995.
PF 12-NOV-1993; 401906.
PR 01-SEP-1989; US-401906.
PR 12-NOV-1991; US-790856.
PR 18-MAY-1993; US-063841.
PR 12-NOV-1993; US-132405.
PA (GETH) GENENTECH INC.
PI Ammann AJ, Rudman CG;
DR WPI: 95-169610/22.
PT Compn. for treating skeletal tissue deficiency - comprising
PT transforming growth factor-beta and an osteogenic cell source in a
PT carrier
PS Claim 3; Column 15-18; 19pp; English.
CC This sequence represents human transforming growth factor-beta 1
CC (TGF-beta 1). This sequence and the sequence for human TGF-beta 3 (see
CC R33598) are claimed within the scope of the invention. The invention is
CC a composition consisting of a TGF-beta protein and an osteogenic cell
CC source (OCS) formulated in an acceptable carrier other than a bone
CC morphogenic cofactor. This composition can be used for the restoration
CC of bone deficiency. This provides for the generation of mature bone only
CC where it is required, without the inclusion of a specific bone-inducing
CC cofactor. This method can be used with any of the 5 human TGF-beta's or
CC with TGF-beta from other species.
SO Sequence 390 AA;

| | | | | |
|-----------------------|--------|---------------------|--------|----------------|
| Query Match | 20.48; | Score 90; | DB 14; | Length 390; |
| Best Local Similarity | 43.88; | Pred. No. 1.30e+01; | | |
| Matches | 14; | Conservative | 6; | Mismatches 12; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

```
db 6 lrllprrllplwllvltgppaagistcktd 37
    | ||:| | | | : | : | | : |
QY 11 LLLISHLGLGWLDLQEAAPADELSSGKMAE 42
```

| | RESULT 13 |
|----|---|
| ID | P81362 standard; protein; 391 AA. |
| AC | P81362; |
| DT | 09-OCT-1990 (first entry) |
| DE | Human transforming growth factor-beta 1. |
| KW | Transforming growth factor-beta 1; tumour treatment. |
| OS | Homo sapiens. |
| FH | Key |
| FT | region |
| FT | 8..21 |
| FT | /label-signal peptide |
| FT | 280..391 |
| FT | /label-mature protein |
| PN | EP-293785-A. |
| PD | 07-DEC-1988. |
| PF | 27-MAY-1988; 108528. |
| PR | 29-MAY-1987; US-055662. |
| PR | 25-JAN-1988; US-147842. |
| PA | (ONCO-) Oncogen. |
| PI | Purchio AF, Gentry L, Twardzik D; |
| DR | WPI; 88-347488/49. |
| DR | N-PSDB; N81085. |
| PT | Prodn. of simian transforming growth factor beta-1 - |
| PT | by culturing transfected eucaryotic cells, and new precursor |
| PT | proteins, useful for treating tumours. |
| PS | Claim 15; page 38; 74pp; English. |
| CC | The mature protein has 100% homology with mature simian TGF-beta 1, which |
| CC | also has almost total homology in the precursor regions, so simian |
| CC | TGF-beta 1 has equivalent biological activity, and is useful in the |
| CC | treatment of tumours. The protein |
| CC | sequence 391 AA; |

| | | | | |
|--------------------------|--------|---------------------|-----------|-------------|
| Query Match | 20.4%; | Score 90; | DB 1; | Length 391; |
| Best Local Similarity | 43.8%; | Pred. No. 1.30e+01; | | |
| Matches 14; Conservative | 6; | Mismatches 12; | Indels 0; | Gaps 0; |

```
Db      6 lrllp1lplwlvltpgppaglstckld 37
```

| QY | 11 | LLLSLHLGLGWLDLQEPAPADELSGKMAE | 42 |
|--------|--|--|----|
| RESULT | 14 | | |
| ID | R83054 | standard; Protein; 391 AA. | |
| AC | R83054; | | |
| DT | 25-JUN-1996 | (first entry) | |
| DE | Transforming growth factor-beta 1. | | |
| KW | macrophage inducible nitric oxide synthase; iNOS; constitutive NOS; | | |
| KW | interleukin-1-beta; transforming growth factor-beta; TGF-beta; IL1-beta; | | |
| KW | nitric oxide production; hypotension; inflammation; septic shock; | | |
| KW | treatment. | | |
| OS | Mammalian sp. | | |
| FT | Key | Location/Qualifiers | |
| FT | protein | 279..391 | |
| FT | | /note-"represents the mature active TGF beta-1 mol." | |
| PN | W09526745-A1. | | |
| PD | 12-OCT-1995. | | |
| PF | 05-APR-1994; U03705. | | |
| PR | 05-APR-1994; WO-U03705. | | |
| PA | (HARD) HARVARD-COLLEGE. | | |
| PI | Lee M, Perrella MA; | | |
| DR | WPI: 95-358443/46. | | |
| DR | N-PSDB; T05876. | | |
| PT | Treatment of hypotension, esp. in septic shock - by administering | | |
| PT | transforming growth factor-beta e.g. to inhibit inducible nitric | | |
| PT | oxide synthase gene transcription | | |
| PS | Disclosure; Fig 17; 52pp; English. | | |
| CC | Transforming growth factor-beta 1 (TGF-beta 1) has been found to inhibit | | |
| CC | inducible nitric oxide synthase (iNOS) gene transcription, esp. in | | |
| CC | interleukin-1-beta (IL1-beta) stimulated rat smooth muscle cells, and at | | |
| CC | a dose which does not inhibit constitutive NOS. TGF-beta 1 or 2 (R83055) | | |
| CC | or their active fragments (esp. derived from the carboxy-terminal 112 | | |
| CC | amino acids), can be used in the treatment of hypotension, such as that | | |
| CC | associated with severe inflammation or septic shock. | | |
| SQ | Sequence 391 AA; | | |

| | | | | |
|-----------------------|--------------|---------------------|----------------|-------------|
| Query Match | 20.4%; | Score 90; | DB 16; | Length 391; |
| Best Local Similarity | 43.8%; | Pred. No. 1.30e+01; | | |
| Matches 14; | Conservative | 6; | Mismatches 12; | Indels 0; |
| | | | Gaps | 0; |

```
Db .      6 lrllplrlplwlvltgppaaagstcktid 37
          | ||:| | | :| | :||:| ||:| :
QY       11 LLLSLHIGLGVLDLQEARPADELSGKMAE 42
```

| | RESULT | 15 |
|----|---|--|
| ID | W78785 | standard; Protein; 390 AA. |
| AC | W78785; | |
| DT | 21-DEC-1998 | (first entry) |
| DE | Human pre-transforming growth factor-beta 1. | |
| KW | Transforming growth factor-beta 1; TGF-beta 1; human. | |
| OS | Homo sapiens. | |
| FH | Key | |
| FT | Domain | |
| FT | | Location/Qualifiers |
| FT | Protein | 8..23 |
| FT | | /note= "hydrophobic domain" |
| FT | | 279..390 |
| FT | | /label= Mat_protein |
| FT | | 82..84 |
| FT | Modified_site | /note= "Asn is N-glycosylated" |
| FT | | 136..138 |
| FT | Modified_site | /note= "Asn is N-glycosylated" |
| FT | | 176..178 |
| FT | Modified_site | /note= "Asn is N-glycosylated" |
| FT | | 277..278 |
| FT | Cleavage_site | /note= "cleavage site for release of TGF-beta 1" |
| PN | US5801231-A. | |
| PD | 01-SEP-1998. | |
| PF | 22-MAR-1985; | 715142. |
| PR | 13-MAR-1987; | US-025423. |
| PR | 22-MAR-1985; | US-715142. |
| PR | 04-AUG-1989; | US-389929. |
| PR | 04-MAR-1992; | US-845893. |

OM of: US-09-030-606-223 to: PIR_60:* out_format : pfs
Date: Sep 25, 1999 6:28 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

```

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171617_29825/app_query.fasta.1
-DB_PIR_60-QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

```

Search information block:

Query: US-09-030-606-223

Query length: 383

Database: PIR_60:*

Database sequences: 122810

Database length: 40065486

Search time (sec): 182.540000

| Sequence | Strd | Orig | ZScore | EScore | len | Documentation |
|-------------|------|-------|--------|--------|------|----------------------------------|
| score_list: | | | | | | |
| ptr2:S64779 | + | 68.00 | 140.35 | 0.7125 | 586 | probable membrane protein YLL02 |
| ptr2:S72642 | + | 66.00 | 140.79 | 1.31 | 301 | probable ABC-type transport prd |
| ptr2:JC5406 | - | 64.50 | 141.20 | 2.07 | 181 | brain and muscle Ah receptor nu |
| ptr2:JC5407 | - | 64.50 | 138.37 | 2.08 | 258 | brain and muscle Ah receptor nu |
| ptr2:PC4288 | - | 64.50 | 137.98 | 2.09 | 271 | brain and muscle Ah receptor nu |
| ptr2:B71538 | + | 64.50 | 134.59 | 2.11 | 414 | probable hypothetical protein c |
| ptr2:S07744 | + | 64.50 | 132.04 | 2.13 | 570 | NADH dehydrogenase (ubiquinone) |
| ptr2:JC5404 | - | 64.50 | 131.86 | 2.13 | 583 | brain and muscle Ah receptor nu |
| ptr2:JC5405 | - | 64.50 | 131.29 | 2.13 | 626 | brain and muscle Ah receptor nu |
| ptr2:S26985 | + | 64.50 | 127.39 | 2.16 | 1021 | probable DNA-directed DNA poly |
| ptr2:S01500 | + | 64.00 | 134.67 | 2.46 | 352 | NADH dehydrogenase (ubiquinone) |
| ptr2:JC4897 | - | 62.50 | 126.42 | 3.98 | 626 | Arnt-like PAS protein, Arnt3 - m |
| ptr2:JC4897 | + | 61.50 | 121.98 | 5.47 | 805 | leptin receptor, Ob-Rb - rat |
| ptr2:JC4797 | + | 61.50 | 121.14 | 5.48 | 894 | leptin receptor precursor - rat |
| ptr2:PC4184 | + | 61.50 | 119.05 | 5.52 | 1162 | leptin receptor, Ob-Rb - rat |
| ptr2:S44840 | - | 61.00 | 135.70 | 6.11 | 124 | KO2D10.2 protein - Caenorhabdit |
| ptr1:WMTGMG | - | 61.00 | 115.23 | 6.50 | 1608 | 183k protein - tobacco mild gr |
| ptr2:D64108 | - | 60.00 | 122.32 | 8.64 | 488 | gluconate transport protein hom |
| ptr2:T02568 | - | 60.00 | 116.68 | 8.79 | 989 | hypothetical protein T16B24.2 |
| ptr2:C71637 | - | 59.50 | 124.88 | 9.98 | 304 | rod shape-determining protein m |
| ptr2:SS7058 | - | 59.50 | 114.46 | 10.30 | 1121 | probable membrane protein YJR0 |
| ptr2:C64432 | + | 59.00 | 119.12 | 11.83 | 537 | hypothetical protein Mj1060 - M |
| ptr2:D71690 | + | 58.50 | 121.39 | 13.68 | 347 | hypothetical protein RP338 - R |
| ptr2:A65163 | - | 58.50 | 121.16 | 13.69 | 357 | lipopolysaccharide 1,2'-n- acety |
| ptr2:S27175 | + | 58.50 | 119.51 | 13.76 | 439 | preprotein translocase secy - S |
| ptr2:A41943 | + | 58.50 | 113.62 | 14.01 | 918 | vacuolar membrane protein PEP3 |
| ptr2:A70471 | + | 58.00 | 124.11 | 15.81 | 212 | phosphoglycerate mutase 1 - Aqu |
| ptr2:F70127 | - | 58.00 | 123.28 | 15.85 | 235 | probable glucose-6-phosphate 1- |
| ptr1:QOBE8 | + | 58.00 | 102.56 | 16.88 | 3149 | BpLF1 protein - human herpesv |
| ptr2:E70106 | + | 57.50 | 122.48 | 18.50 | 223 | uracil DNA glycosylase (ung) hc |
| ptr2:I39826 | + | 57.50 | 118.12 | 18.75 | 385 | ORF2 - Bacillus firmus |
| ptr2:A33991 | + | 57.50 | 114.09 | 18.98 | 638 | somatotropin receptor precursor |
| ptr2:S04530 | + | 57.50 | 114.09 | 18.98 | 638 | somatotropin receptor, hepatic |
| ptr2:B70001 | + | 57.50 | 113.99 | 18.99 | 646 | ABC transporter (permease) homc |
| ptr2:F71727 | - | 57.50 | 110.44 | 19.19 | 1008 | acriflavin resistance protein |
| ptr1:C5MS | + | 57.50 | 106.36 | 19.43 | 1680 | complement C5 precursor - mous |
| ptr2:G70179 | + | 57.00 | 119.95 | 21.72 | 263 | spermidine/putrescine ABC trans |
| ptr2:D71086 | - | 56.50 | 125.41 | 24.88 | 114 | hypothetical protein PH0953 - H |
| ptr2:H71208 | + | 56.50 | 120.27 | 25.27 | 217 | hypothetical protein PH1936 - H |
| ptr2:G64642 | + | 56.50 | 118.40 | 25.42 | 274 | conserved hypothetical integral |
| ptr2:A45033 | - | 56.50 | 110.63 | 26.02 | 725 | myelin transcription factor 1 - |
| ptr2:A71704 | - | 56.50 | 110.01 | 26.07 | 784 | ATP-dependent proteinase LA (ld |

```

pir1:C5HU          + 56.50 103.94 26.56 1676 1 complement C5 precursor - h
pir2:E71495        + 56.00 120.19 29.44 188 1 probable peptidoglycan-assoc
pir2:B70693         + 56.00 117.95 29.65 249 1 probable echa16 protein - My

seq_name: pir2:S64779

seq_documentation_block:
Probable membrane protein YLL028w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L0939
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 14-Nov-1997
C:Accession: S64779
R:Duesterhoeft, A.; Floeth, M.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64775
A:Accession: S64779
A:Molecule type: DNA
A:Residues: 1-586 <DUE>
A:Cross-references: EMBL:Z73133; NID:g1360209; PID:e245461; PID:g1360210; MIPS:YLL028
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 12L
C:Keywords: transmembrane protein
F:146-162/Domain: transmembrane #status predicted <TM1>
F:179-195/Domain: transmembrane #status predicted <TM2>
F:214-230/Domain: transmembrane #status predicted <TM3>
F:240-256/Domain: transmembrane #status predicted <TM4>
F:273-289/Domain: transmembrane #status predicted <TM5>
F:306-322/Domain: transmembrane #status predicted <TM6>
F:369-385/Domain: transmembrane #status predicted <TM7>
F:389-405/Domain: transmembrane #status predicted <TM8>
F:416-432/Domain: transmembrane #status predicted <TM9>
F:457-473/Domain: transmembrane #status predicted <TM10>
F:489-505/Domain: transmembrane #status predicted <TM11>
F:553-569/Domain: transmembrane #status predicted <TM12>

alignment_scores:
Quality: 68.00 Length: 92
Ratio: 1.581 Gaps: 5
Percent Similarity: 46.739 Percent Identity: 26.087

alignment_block:
US-09-030-606-223 x S64779 ..

Align seg 1/1 to: S64779 from: 1 to: 586

91 GGGGCAATTCCTTACATGTGCTGACACAGATTAAATGCTGTGCCAAAA 140
||| ::||| ||||| ::|||
414 GlyGluLeuProTyrIleAlaLeuIleIleGlyMetMetValCysAlaAl 430

141 TTTTGATTTTATTGGAGACTTCTTATCAAAAGTAATGCTGCCCAAAGA 190
||||:||||:||||:||||: ||| ||| |||
430 aPheIleTyrPylMetAspAsnAspTyrLeuLysArgCys....ArgA 445

191 AGTCTAAGCAATTAGTA..... 207
::: ::||| |||
445 IalysGlyLysLeuValProGluAlaArgLeuTyrAlaMetValIleAla 461

208 .....GTGTCCCM...TCACGTGTTGGAGTGTGCTATTCTAAAGAT 248
||||| ||| ::||| |||
462 GlyThrValPheProIleGlyIleLeuTyr..... 471

249 TTGATTTCTCGCAATGACAAATTAT.....ATTTTAA 280
::: |||: |||
472 ...PheCysTyrPThrGlyTyrTyrProHisLysIleHisTyrMetValP 487

281 CTTTGTGGGGGAANAGTTATAGA 306
||||| ::| |||||
487 roThrValGlyGlyAlaPheIleGly 495

seq_name: pir2:S72642

```


seq_documentation_block:

probable ABC-type transpore protein xynB - Thermoanaerobacterium thermosulfurigenes
C;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 24-Jul-1998
C;Accession: S72642
R;Matuschek, M.; Sahm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes F
A;Reference number: S72640
A;Accession: S72642
A;Molecule type: DNA
A;Residues: 1-301 <MAT>
A;Cross-references: EMBL:U50951
A;Experimental source: strain EML
C;Genetics:
A;Gene: xynB
A;Start codon: GTG

alignment_scores:

| Quality: | 66.00 | Length: | 108 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.200 | Gaps: | 4 |
| Percent Similarity: | 50.926 | Percent Identity: | 24.074 |

alignment_block:

US-09-030-606-223 x S72642 ..

Align seg 1/1 to: S72642 from: 1 to: 301

```
70 TGGTCATTTAATWRRRTKGGGCGCATTCCTTACATGCTGTGACAAG 119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
121 TyrSerMetTyrPheThrSergIleuProTyrTyrIleuIleuIleu 137
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
120 ATTAATAATGCTGTGCCAAATTTGTATTATTGAGACTTCTTATC 169
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
137 GASPLEUHSLEUMETASNPHELEUVALTYR..... 148
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
170 AAAAGTAATGCTGCCAAGAAGTCTAAGGAATGAGTGTCCCMTC 219
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
149 ..... 152
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
220 CTGTGTT.....TGGAGTGTCTATTCTAAAGATTTGATTTCCGTGAA 263
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
153 ILeValSerAlaTrpAsnIleMet.....ValValArgSerTyrI 166
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
264 TGACAATTATATTCTTGGTGGGGGGAANAGTATAGACACACAGT 313
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
166 eAspGlyLeuProAlaSerLeuValGluSerAlaLysIleAspGlyAla 183
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
314 CTTCACCTTCTGATACCTGTAATATATCTTT.....ATGCACCTT 354
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
183 eGluLeuArgIleIlePheSerIleIlePheProLeuSerValProVal 199
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
355 GTTTGACCATTAAGCTATATGTT 378
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
200 LeuAlaThrIleThrLeuPheVal 207
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

seq_name: p1r2:JC5406

seq_documentation_block:

brain and muscle Ah receptor nuclear translocator-like protein, BMAL1c - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 10-Sep-1997
C;Accession: JC5406
R;Ikeda, M.; Nomura, M.
Biochem. Biophys. Res. Commun. 233, 258-264, 1997
A;Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/R
site usage.
A;Reference number: JC5404; MUID:97289529
A;Accession: JC5406
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-181 <IKE>

A;Cross-references: DDBJ:AB000813; NID:g2094736; PID:d1020726; PID:g2094737

C;Comment: This protein plays a role in the control of central nervous system develop

alignment_scores:

| Quality: | 64.50 | Length: | 76 |
|--|--|-----------------------------------|--------|
| Ratio: <td>1.697<td>Gaps:<td>1</td></td></td> | 1.697 <td>Gaps:<td>1</td></td> | Gaps: <td>1</td> | 1 |
| Percent Similarity: <td>50.000<td>Percent Identity:<td>23.684</td></td></td> | 50.000 <td>Percent Identity:<td>23.684</td></td> | Percent Identity: <td>23.684</td> | 23.684 |

alignment_block:

US-09-030-606-223/rev x JC5406 ..

Align seg 1/1 to: JC5406 from: 1 to: 181

```
292 CCCCACCAAGTTAAATATATATGTCATTCAGGAATCAAAATCTTT 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
14 ProHISGlyArgLeuGluTyrThrGluHISGlnGlyArgIleLysAsnAl 30
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
242 TAGAATAGCACACATCCCAACAAGTGAKGGGAACACTACTAATTCCTTAGA 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
30 aArgGluAlaHisSerGln..... 36
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
192 CTTCCTTGGCAGCATTACTTTTGATAGAACTCTCCAAATTAATACAA 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
37 .....IleGluLysArgArgArgAspLysMetAsn 46
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
142 AATTTGGCAGACATTTTAACTTGTCAAGACAATGTAGGAATGCC 93
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
47 SerPheIleAspGluLeuAlaSerLeuValProThrCysAsnAlaMetSe 63
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
92 CCMAAYAYWATTAATTGACCATTAAT 65
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
63 rArgLysLeuAspLysLeuThrValLeu 72
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

seq_name: p1r2:JC5407

seq_documentation_block:

brain and muscle Ah receptor nuclear translocator-like protein, BMAL1c - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 10-Sep-1997
C;Accession: JC5407
R;Ikeda, M.; Nomura, M.
Biochem. Biophys. Res. Commun. 233, 258-264, 1997
A;Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix
site usage.
A;Reference number: JC5404; MUID:97289529
A;Accession: JC5407
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-258 <IKE>
A;Cross-references: DDBJ:AB000815; NID:g2094740; PID:d1020728; PID:g2094741
C;Comment: This protein plays a role in the control of central nervous system develop

alignment_scores:

| Quality: | 64.50 | Length: | 76 |
|--|--|-----------------------------------|--------|
| Ratio: <td>1.697<td>Gaps:<td>1</td></td></td> | 1.697 <td>Gaps:<td>1</td></td> | Gaps: <td>1</td> | 1 |
| Percent Similarity: <td>50.000<td>Percent Identity:<td>23.684</td></td></td> | 50.000 <td>Percent Identity:<td>23.684</td></td> | Percent Identity: <td>23.684</td> | 23.684 |

alignment_block:

US-09-030-606-223/rev x JC5407 ..

Align seg 1/1 to: JC5407 from: 1 to: 258

```
292 CCCCACCAAGTTAAATATATATGTCATTCAGGAATCAAAATCTTT 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
14 ProHISGlyArgLeuGluTyrThrGluHISGlnGlyArgIleLysAsnAl 30
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
242 TAGAATAGCACACATCCCAACAAGTGAKGGGAACACTACTAATTCCTTAGA 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
30 aArgGluAlaHisSerGln..... 36
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
192 CTTCCTTGGCAGCATTACTTTTGATAGAACTCTCCAAATTAATACAA 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```


Quality: 64.50 Length: 119
Ratio: 1.093 Gaps: 7
Percent Similarity: 49.580 Percent Identity: 25.210

alignment_block:

US-09-030-606-223 x S07744 ..

Align seg 1/1 to: S07744 from: 1 to: 570

```
105 CATGTGCTTGACAGATTAAATG...TCTGTGCCAAA..... 140
||| ::::::::::::::|::|::|::|::|
203 HisLeuTyrSerGluMetLysLeuGlySerThrProGlnIleAsnSerTr 219
141 .....TTTGTATTTTATTGTGAGACTTCTATCAAAA.... 173
||| ::::::::::::::|::|::|::|::|
219 PAsnLeuIleSerPheCysLeuLeuPheAlaIaIaPheValLysSerAlaG 236
174 .....GTAATGCTGCCAAGAGAGCTTAAGGAATTAGTACT 209
||| ::::::::::::::|::|::|::|::|
236 InPheGlyPheHisValThrLeuProAspSerMet.....GluAlaPro 250
210 GTTCCCMTCACCTGTTTG..... 229
||| ::::::::::::::|::|::|::|::|
251 ValProAlaSerAlaLeuIleHisSerAlaThrLeuValSerAlaGlyVa 267
230 GTGTGCTATTCTAAAGATTGTTGATTCCTGGAATGACAATTATATTGTA 279
||| ::::::::::::::|::|::|::|::|
267 LpheLeuIleMetArgPheTyrProIleLeuGluLeuSerLeuTyrPheL 284
280 AC.....TTTGTGGGGGAAN 296
::| ::::::::::::::|::|::|::|::|
284 ysLeuValThrAlaLeuValGlyAlaLeuThrAlaLeuAlaGlyGlyLeu 300
297 AGTTATAGACACAGTCTTCACTTCTGATCTGTAATTAATCTTTTA 346
||| ::::::::::::::|::|::|::|::|
301 Ser.....AlaValPheGlnThrAspLeuLysLysIleLeuAlaTy 314
347 TTGCACT 353
| |||
314 rSerThr 316
```

seq_name: plr2:JC5404

seq_documentation_block:

brain and muscle Ah receptor nuclear translocator-like protein, BMAL1a - human
C/Species: Homo sapiens (man)
C/Date: 10-Jun-1997 #sequence_revision 10-Jun-1997 #text_change 18-Jul-1997
C/Accession: JC5404
R/Ikeda, M.; Nomura, M.
Biochem. Biophys. Res. Commun. 233, 258-264, 1997
A/Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/F
site usage.
A/Reference number: JC5404; MUID:97289529
A/Accession: JC5404
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-583 <IKE>
A/Cross-references: DDBJ:D89722
C/Comment: This protein plays a role in the control of central nervous system developmen
C/Genetics:
A/Introns: 4/2; 180/1; 207/2; 231/1; 245/2; 318/1; 349/1; 399/2

alignment_scores:

Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 1
Percent Similarity: 50.000 Percent Identity: 23.684

alignment_block:

US-09-030-606-223/rev x JC5404 ..

Align seg 1/1 to: JC5404 from: 1 to: 583

```
292 CCCCCACCAAGTTAAATATATATGTCATCCAGGAATCAAAATCTTT 243
||| ::::::::::::::|::|::|::|::|
14 ProHisGlyArgLeuGluTyrThrGluHisGlnGlyArgIleLysAsnAl 30
```

242 TAGAATAGCACACTCCAAACAAGTGAKGGGAACACTACTAATTCCTTGA 193

||| ::::::::::::::|::|::|::|::|
30 aArgGluAlaHisSerGln..... 36

192 CTTCCCTTGGCAGCATTACTTTTGATAGAAGTCTCCAAATTAATACAA 143

37IleGluLysArgArgArgAspLysMetAsn 46

142 AATTTGGCACACACATTTTAATCTGTCAAGACAAATGTAAGCAATGCC 93

47 SerPheIleAspGluLeuAlaSerLeuValProThrCysAsnAlaMetSe 63

92 CCAMAAYAYWATTAATTGACCATTAAT 65

::| ::::::::::::::|::|::|::|::|
63 rArgLysLeuAspLysLeuThrValLeu 72

seq_name: plr2:JC5405

seq_documentation_block:

brain and muscle Ah receptor nuclear translocator-like protein, BMAL1b - human
C/Species: Homo sapiens (man)
C/Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 10-Sep-1997
C/Accession: JC5405
R/Ikeda, M.; Nomura, M.
Biochem. Biophys. Res. Commun. 233, 258-264, 1997
A/Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix
site usage.
A/Reference number: JC5404; MUID:97289529
A/Accession: JC5405
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-626 <IKE>
A/Cross-references: DDBJ:AB000812; NID:g2094734; PID:d1020725; PID:g2094735
C/Comment: This protein plays a role in the control of central nervous system develop

alignment_scores:

Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 1
Percent Similarity: 50.000 Percent Identity: 23.684

alignment_block:

US-09-030-606-223/rev x JC5405 ..

Align seg 1/1 to: JC5405 from: 1 to: 626

292 CCCCCACCAAGTTAAATATATATGTCATCCAGGAATCAAAATCTTT 243

||| ::::::::::::::|::|::|::|::|
57 ProHisGlyArgLeuGluTyrThrGluHisGlnGlyArgIleLysAsnAl 73

242 TAGAATAGCACACTCCAAACAAGTGAKGGGAACACTACTAATTCCTTGA 193

||| ::::::::::::::|::|::|::|::|
73 aArgGluAlaHisSerGln..... 79

192 CTTCCCTTGGCAGCATTACTTTTGATAGAAGTCTCCAAATTAATACAA 143

80IleGluLysArgArgArgAspLysMetAsn 89

142 AATTTGGCACACACATTTTAATCTGTCAAGACAAATGTAAGCAATGCC 93

90 SerPheIleAspGluLeuAlaSerLeuValProThrCysAsnAlaMetSe 106

92 CCAMAAYAYWATTAATTGACCATTAAT 65

::| ::::::::::::::|::|::|::|::|
106 rArgLysLeuAspLysLeuThrValLeu 115

seq_name: plr2:S26985

seq_documentation_block:

probable DNA-directed DNA polymerase (EC 2.7.7.7) - Neurospora crassa mitochondrion plasmid
C/Species: mitochondrion Neurospora crassa
C/Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 13-Sep-1998
C/Accession: S26985
R/Court, D.A.; Bertrand, H.
Curr. Genet. 22, 385-397, 1992
A/Title: Genetic organization and structural features of maranhar, a senescence-inducing
A/Reference number: S26984; MUID:93046810
A/Accession: S26985
A/Molecule type: DNA
A/Residues: 1-1021 <COU>
A/Cross-references: EMBL:X55361; NID:g3023; PID:e246763; PID:g578156
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SGC3
C/Superfamily: Neurospora crassa mitochondrion plasmid probable DNA-directed DNA polymerase
C/Keywords: DNA binding; mitochondrion; nucleotidyltransferase

```
alignment_scores:      .  
    Quality:          64.50  
     Ratio:           1.613  
Percent Similarity:   55.556  
  
                       Length:       72  
                        Gaps:         3  
Percent Identity:    27.778
```

```
alignment_block:
US-09-030-606-223 x S26985 ..
```

Align seg 1/1 to: S26985 from: 1 to: 1021

```

138 AAATTTGTATTTATTTGGAGACTTCTTATCAAAGTAATGCTGCCAA 187
      |||||  |||  ::|||::: ||:::~::~: ::
383 LyspHeaspIle.....AspHeileIleLysIleLeuValGIngl 396
      :  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
188 GGAAGTCTAAGGAATTAGTAGTGTTCCMTCACCTGTTTGGAGTGTGCTA 237
      :  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
396 upHeval.....ValGluLysIleI 403
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
238 TTCTAAAGATTTTGTATTTCCCTGGAATGACAATTATATTTAACTT... 284
      ||  |||||:::||||:~::~: ~::~: ~::~: ~::~: ~::~: ~::~:
403 IeSerLysAspLeuAspIleLeuSerIleLysIleSerTyrLysPheGlu 419
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
285 .....GGTGGGGAANAAGTTATAGCACCACAGTCTTCACTTCTGA 325
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
420 ProLysLysLysGlyGlyLysAlaGluArgHisThrIleThrIleAlaAs 436
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
326 TACTTGTAATTAATC 341
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
436 pSerCysArgLeuLeu 441
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

```

seq_name: plr2:S01500

seq_documentation_block:

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - sea urchin (*Strongylocentrotus purpuratus*)
C/Species: mitochondrion *Strongylocentrotus purpuratus* (purple urchin)
C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 08-Sep-1997
C/Accession: S01500
R/Jacobs, H.T.; Elliott, D.J.; Math, V.B.; Farquharson, A.

J. Mol. Biol. 202, 185-217, 1988

A;Title: Nucleotide sequence and gene organization of sea urchin mitochondrial DNA.
A;Reference number: S01499; MUID:89011951

A;Accession: S01500

A;Molecule type: DNA

A;ResIdues: 1-352 <JAC>

A:Cross-references: EMBL:X12631; NID:g296545; PID:q13666

A; Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 136-Met

C:Genetics:

A;Gene: nd2

A; Genome: mitochondrion

A;Genetic code: SGC8

C;Superfamily: NADH dehydrogenase (ubiquinone) chain 2

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 64.00 | Length: 142 |
| Ratio: | 0.889 | Gaps: 8 |
| Percent Similarity: | 50.704 | Percent Identity: 22.535 |

alignment_block:

US-09-030-606-223 x 501500

Align seg 1/1 to: S01500 from: 1 to: 352

```

21  ACAATTCCTTCATTCAGAAAAATATCTTAGGGACTGATATTGGTAATTAT 70
    |||||  :::  :::  |||  ::|||::|||:::||||
232  ThrIleSerGlnLeuSerProIleSerValAlaIleuValLeuIleuValMe 248
71  GGTCAATTAAATWRTTRTKTGGGCATTCCCT..... 102
    |:::||||  |||::  |||
248  tIeuSerIeu.....GlycylIeu.ProProIeuThrglyPheIle 261
103  .....TACATTGCTCT 113
    ::|||::|
262  leuIysPheIhrSerLeuTyrPheIeuValAlaIasnaSnPheIleIle 278
114  GACAAGATTAAATGTCTG.....TGCCAAATTTTGTAATTTATTTGG 157
    |::  ::  ::  |||:::  ||||||
278  uSerSerIleuMetIleIleGlyAsnIeuGlnAspTyrPhePheTyrIeuA 295
158  AGACTTCCTTAACAAAGAATATGCTGCCAAGAAGTCTAAGAATTAGTA 207
    ::  |||::  ::  ::|||
295  rglIeSerPhe.....AsnThrSerIeuPhe 303
    ::|||::|
208  GTGTTCCTCCMTCACCTGTTGGAGTGTCTATTCTAAAGATTTGATTTC 257
    ::||||||::  ::  |||
304  leuPheProGlnHisIleIleSer.....SerAlaIse 314
    ||
258  CTGGAATGACAATATATTTTAAC..TTGGTGGGGGAANAAGTTATAG 304
    |||:::  ::  |||:::  |||::  :::  :::
314  rTrpArgAsnSerThrIleIleIleSerProIleuAlaProIysAlaTyrIeuS 331
305  GACCACAGTCTTCACCTCTGTGACTTGTAAATTAACTTTATTCGACCT 354
    ::  |||:::  |||  |||  ::|||::
331  erSerValSerThrValIeu.....SerThrIleuAlaIle 342
355  GTTTTGACCATTAAGCTATATGTT 378
    |||||::  |||||::
343  ProIeuThrIeuProIeuTyrIle 350

```

seq_name: p1r2:JE0270

seq_documentation_block:

Arnt-like PAS protein, Arnt3 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 05-Feb-1999 #sequence revision 05-Feb-1999 #text change 05-Feb-1999

C:\Accesslon: JE0270

R; Takahata, S.; Sogawa, K.; Kobayashi, A.; Ema, M.; Mimura, J.; Ozaki, N.; Fujii-Kuri

Biochem. Biophys. Res. Commun. 248, 789-794, 1998

A;Title: Transcriptionally active heterodimer formation of an Arnt-like PAS protein,Ar

A; Reference number: JE0270

A;Accession: JE0270

A; molecule type: mRNA

A;Residues: 1-626 <TAK>

A;Cross-references: DDBJ:AB014494

alignment_scores:

| | | | |
|----------|-------|---------|----|
| Quality: | 62.50 | Length: | 76 |
|----------|-------|---------|----|

Ratio: 1.689 Gaps: 1

Percent Similarity: 48.684 Percent Identity: 23.684

alignment_block:

US-09-030-606-223/rev x JE0270 . .

Align seg 1/1 to: JE0270 from: 1 to: 626

seq_name: p1r2:PC4184

seq_documentation_block:

leptin receptor, Ob-Rb - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: J04895; PC4184
R:Takaya, K.; Ogawa, Y.; Isse, N.; Okazaki, T.; Satoh, N.; Masuzaki, H.; Mori, K.; Tamura
Biochem. Biophys. Res. Commun. 225, 75-83, 1996
A:Title: Molecular cloning of rat leptin receptor isoform complementary DNAs-identification
A:Reference number: J04895; MUID:96332408
A:Accession: J04895
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1162 <TAK>
A:Cross-references: DDBJ:D85558; NID:g1526441; PID:d1013515; PID:g1526442
R:Iida, M.; Murakami, T.; Ishida, K.; Mizuno, A.; Kuwajima, M.; Shima, K.
Biochem. Biophys. Res. Commun. 224, 597-604, 1996
A:Title: Substitution at codon 269 (glutamine-proline) of the leptin receptor (OB-R) cDN
A:Reference number: PC4184; MUID:96295531
A:Accession: PC4184
A:Molecule type: mRNA
A:Residues: 840-1162 <IID>
A:Cross-references: DDBJ:D84550
C:Comment: This receptor is obese-phenotype-linked mutant, found in the Zucker (fa/fa) m
C:Keywords: appetite; transmembrane protein
F:840-860/Domain: transmembrane #status predicted <TMM>
F:861-1162/Domain: intracellular #status predicted <INT>

alignment_scores:

| Quality: | 61.50 | Length: | 111 |
|---------------------|--------|-------------------|--------|
| Ratio: | 0.992 | Gaps: | 8 |
| Percent Similarity: | 55.856 | Percent Identity: | 31.532 |

alignment_block:

US-09-030-606-223 x PC4184

Align seg 1/1 to: PC4184 from: 1 to: 1162

```
17 AAAAACAATCT.....TCATTCAGAAAAATATCTTA..... 49
   |||||:|||||   ||| |||:|   :|||:|
667 LysAsnAspSerLeuCySerValArgTyrValValLysHisArgTh 683
   50 .....GGGACT.....GATATTGGTAATATGCTCAATTGA 80
      |||||   |||||   |||:|||||   :|||:|
683 rAlaHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 699
   81 ATWRTTRTKGGGCATTTCCCTTACAT.....TGCTCTTGA 115
      |||:|   |||:|   |||   |||   :|||:|
700 ThrPheLeuTrpAlaGluSerAlaHisThrValThrValLeuAlaIleAs 716
   116 CAAGATTAATAATGCTCTGTGCCAAATTTTGTATTTATTTGGAGACTTCT 165
      :|||:|   |||:|   :|||:|   :|||:|   :|||:|
716 nSerIleGlyAlaSerLeuValAsnPhaAsnLeuThrPheSerTrpProm 733
   166 TATCAAAAGTAATGCTGCCAAAGGAAGCTTAAGGAATTAGTGTCTCC 215
      :|||:|   |||:|   :|||:|   :|||:|   :|||:|
733 etSerLysVal.....AsnAlaValGlnSerLeuSerAlaTyrPro 746
   216 MTCACCTGTGTTGGAGTGTGCTATTCCTAAAGATTGATTCCTGCAATG 265
      :|||:|   :|||:|   :|||:|   :|||:|   :|||:|
747 LeuSerSer...SerCysValIleLeu.SerTrpThrLeuSerProAsnA 762
   266 ACAATTATATTTTAACCTTTGGTGGGGGAA 294
      ||   :|||:|   |||:|   |||:|
762 sPtyrSerLeuLeuTyrLeuValIleGlu 771
```


This Page Blank (uspto)

OM of: US-09-030-606-223 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171618_29883/app_query.fasta.1
-DB=SwissProt_37 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-223
Query length: 383
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|---------------------------|-----------|--------|--------|------|-------------------------------|
| SwissProt_37:NUSM_PARTE + | 64.50 | 129.50 | 2.08 | 570 | P15584 parametrium tetraureli |
| SwissProt_37:BMAL_HUMAN - | 64.50 | 129.31 | 2.08 | 583 | P00327 homo sapiens (human). |
| SwissProt_37:DPOK_NEUCR + | 64.50 | 124.76 | 2.13 | 1021 | P33537 neurospora crassa. f |
| SwissProt_37:N2M_STRPU + | 64.00 | 132.24 | 2.37 | 352 | P15549 strongylocentrotus pu |
| SwissProt_37:YMO2_CABEL - | 61.00 | 133.66 | 5.60 | 124 | P34493 caenorhabditis elegan |
| SwissProt_37:RRPO_TMGMV - | 61.00 | 112.82 | 6.26 | 1608 | P18339 tobacco mild green m |
| SwissProt_37:YU09_YEAST - | 59.50 | 112.22 | 9.69 | 1121 | P47107 saccharomyces cerevi |
| SwissProt_37:RFAK_ECOLI - | 58.50 | 117.49 | 12.59 | 439 | P31159 synecchococcus sp. (st |
| SwissProt_37:SECY_SYNP7 + | 58.50 | 111.49 | 13.00 | 918 | P27801 saccharomyces cerevis |
| SwissProt_37:PEP3_YEAST + | 58.00 | 115.18 | 14.74 | 504 | P27664 helicoverpa armigera |
| SwissProt_37:C6B2_HELM + | 58.00 | 100.29 | 15.94 | 3149 | P03186 epstein-barr virus (st |
| SwissProt_37:TEGU_EBV + | 57.50 | 120.64 | 16.55 | 223 | P051082 borrelia burgdorferi |
| SwissProt_37:YCT2_BACFI + | 57.50 | 116.20 | 16.94 | 385 | P04454 bacillus firmus. hypo |
| SwissProt_37:GHR_HUMAN + | 57.50 | 112.09 | 17.31 | 638 | P10912 homo sapiens (human). |
| SwissProt_37:CO5_MOUSE + | 57.00 | 104.22 | 18.05 | 1680 | P06684 mus musculus (mouse). |
| SwissProt_37:YC10_KLEPN + | 57.00 | 114.14 | 19.80 | 429 | P048456 klebsiella pneumoniae |
| SwissProt_37:MYT1_HUMAN - | 56.50 | 108.69 | 23.55 | 725 | P01538 homo sapiens (human). |
| SwissProt_37:CO5_HUMAN + | 56.50 | 101.88 | 24.41 | 1676 | P01031 homo sapiens (human). |
| SwissProt_37:FABL_CHICK - | 56.00 | 121.81 | 25.40 | 125 | P80226 gallus gallus (chicke |
| SwissProt_37:PSD5_HUMAN - | 56.00 | 110.47 | 26.96 | 504 | P16401 homo sapiens (human). |
| SwissProt_37:IMH3_CANAL - | 56.00 | 110.20 | 27.00 | 521 | P00086 candida albicans (yea |
| SwissProt_37:RPOA_SPIOL - | 55.50 | 112.62 | 30.82 | 335 | P06505 spinacia oleracea (sp |
| SwissProt_37:SPAS_SALTY + | 55.50 | 112.12 | 30.90 | 356 | P40702 salmonella typhimuritu |
| SwissProt_37:YC24_CYAPA + | 55.50 | 109.59 | 31.31 | 486 | P48260 cyanophora paradoxa. |
| SwissProt_37:EX01_YEAST - | 55.50 | 106.60 | 31.81 | 702 | P39875 saccharomyces cerevis |
| SwissProt_37:CN3B_HUMAN - | 55.50 | 102.86 | 32.45 | 1112 | P13370 homo sapiens (human) |
| SwissProt_37:RPA1_MOUSE - | 55.50 | 99.33 | 33.06 | 1717 | P035134 mus musculus (mouse) |
| SwissProt_37:YF71_CABEL + | 55.00 | 115.80 | 35.03 | 196 | P09220 caenorhabditis elegan |
| SwissProt_37:YRBE_HAEIN + | 55.00 | 113.47 | 35.46 | 261 | P45030 haemophilus influenza |
| SwissProt_37:CYST_SYNY3 + | 55.00 | 112.72 | 35.60 | 286 | P01895 synecchocystis sp. (st |
| SwissProt_37:US20_HCMVA + | 55.00 | 111.27 | 35.87 | 342 | P09724 human cytomegalovirus |
| SwissProt_37:US20_HCMVT + | 55.00 | 111.27 | 35.87 | 342 | P003307 human cytomegalovirus |
| SwissProt_37:YNOB_PSEAE - | 55.00 | 106.54 | 36.78 | 612 | P51484 pseudomonas aeruginos |
| SwissProt_37:MCM4_MOUSE - | 55.00 | 103.75 | 37.33 | 862 | P49717 mus musculus (mouse). |
| SwissProt_37:COBL_METVA + | 54.50 | 114.02 | 40.87 | 211 | P58917 methanococcus jannasch |
| SwissProt_37:Y184_MYCPN + | 54.50 | 110.66 | 41.60 | 319 | P50290 mycoplasma pneumoniae |
| SwissProt_37:QNEA_AOUAE + | 54.50 | 110.28 | 41.68 | 334 | P67043 aquifex aeolicus. s-d |
| SwissProt_37:RFBX_ECOLI + | 54.50 | 108.52 | 42.07 | 415 | P37746 escherichia coli. put |
| SwissProt_37:FUMH_YEAST - | 54.50 | 107.20 | 42.37 | 488 | P08417 saccharomyces cerevis |
| SwissProt_37:ACM3_BOVIN - | 54.50 | 105.66 | 42.71 | 590 | P41984 bos taurus (bovine). |
| SwissProt_37:G6PC_TOBAC - | 54.50 | 105.62 | 42.72 | 593 | P43793 nicotiana tabacum (cd |

SwissProt_37:UVRC_HAEIN - 54.50 105.40 42.77 609 P44489 haemophilus influe
SwissProt_37:POL2_TBRSV + 54.50 98.89 44.27 1357 P14547 tomato black ring
SwissProt_37:T2D2_YEAST - 54.50 98.59 44.34 1407 P23255 saccharomyces cer

seq_name: SwissProt_37:NUSM_PARTE

seq_documentation_block:
ID NUSM_PARTE STANDARD; PRT; 570 AA.
AC P15584;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (Ec 1.6.5.3).
GN NDS OR NDSH.
OS PARAMECIUM TETRAURELIA.
OG MITOCHONDRION.
OC EUKARYOTA; ALVEOLATA; CILIOPHORA; NASSOPHOREA; PENICULIDA; PARAMECIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STOCK 51.
RX MEDLINE; 90174913.
RA PRITCHARD A.E., SEILHAMER J.J., MAHALINGAM R., SABLE C.L.,
RA VENUTI S.E., CUMINGS D.J.;
RT "Nucleotide sequence of the mitochondrial genome of Parametrium.";
RL NUCLEIC ACIDS RES. 18:173-180(1990).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X15917; G13276; -.
DR PIR; S07744; S07744.
DR PFAM; PF00361; oxidored_g1; 1.
DR PEAM; PF00662; oxidored_g1_N; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 570 AA; 65196 MW; 85B0309F CRC32;

alignment_scores:
Quality: 64.50 Length: 119
Ratio: 1.093 Gaps: 7
Percent Similarity: 49.580 Percent Identity: 25.210

alignment_block:

US-09-030-606-223 x NUSM_PARTE ..

Align seg 1/1 to: NUSM_PARTE from: 1 to: 570

105 CATGCTGTGACAGATTAAATG...TCTGTGCCAAA..... 140
|||:||||:| |||:||||:
203 H1sleutyserglumetylsleuglyserthrproglinileasnsertr 219
141TTTGTAATTATTGAGACTCTTATCAAAA.... 173
|||:||||:| |||:||||:
219 pasnleuileserphecysleuleuphealaalaphelvalysseralag 236
174GTAATGCTGCCAAAGAGCTTAAGAAATTAGTACT 209
|||:||||:| |||:||||:
236 lnphегlyphehisvaltrleuproaspsermet..... Glualapro 250
210 GTTCCCMTCACCTGTTTC..... 229
|||:||||:| |||:||||:
251 Valproalaseralaleuilehisseralathleuvalseralaglyva 267
230 GTGTGCTATCTTAAAGATTGATTCTCGGAATGACAAATATATTTTA 279
|||:||||:| |||:||||:
267 lpheluleulemetargphetyrproilleugluuenserleutyrrhel 284

280 AC.....TTTGTGGGGAAN 296
:::|||||
284 ysleuValThrAlaLeuValGlyAlaLeuThrAlaLeuAlaGlyGlyLeu 300
297 AGTTATAGACACAGCTCTCAGTCTGATCTGTAATTAATCTTTA 346
|||:::|||||:::|||||
301 Ser.....AlaValPheGlnThrAspLeuLysLysIleLeuAlaTy 314
347 TTGCACACT 353
I|||
314 rSerThr 316

seq_name: SwissProt_37:BMAL_HUMAN

seq_documentation_block:

ID BMAL_HUMAN STANDARD; PRT; 583 AA.
AC 000327; 000313; 000314; 000315; 000316; 000317; 099631; 099649;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE BMAL1 PROTEIN (MEMBER OF PAS PROTEIN 3) (MOP3) (BHLH-PAS PROTEIN JAP3).
GN BMAL1 OR ARNTL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS BMAL1 A-F).
RC TISSUE-BRAIN;
RX MEDLINE; 97289529.
RA IKEDA M., NOMURA M.;
RT "cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS protein (BMAL1) and identification of alternatively spliced variants with alternative translation initiation site usage.";
RT BIOCHEM. BIOPHYS. RES. COMMUN. 233:258-264(1997).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM MOP3).
RC TISSUE-FETAL BRAIN;
RX MEDLINE; 97236817.
RA HOGENESCH J.B., CHAN W.K., JACKIM V.H., BROWN R.C., GU Y.-Z.,
RA PRAY-GRANT M., PERDEW G.H., BRADFIELD C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioloxin signaling pathway.";
RT J. BIOL. CHEM. 272:8581-8593(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM BMAL1B).
RA TIAN H., RUSSELL D.W., MCKNIGHT S.L.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER BHLH PROTEIN. INTERACTS WITH HSP90; WITH AHR IN VITRO, BUT NOT IN VIVO.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE ADULT BRAIN, SKELETAL MUSCLE AND HEART.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST SEVEN ISOFORMS; BMAL1A (SHOWN HERE), BMAL1B/JAP3, BMAL1C, BMAL1D, BMAL1E, BMAL1F AND MOP3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D89722; E316533; -
DR EMBL; AB000812; D1020725; -

DR EMBL; AB000813; D1020726; -
DR EMBL; AB000814; D1020727; -
DR EMBL; AB000815; D1020728; -
DR EMBL; AB000816; D1020729; -
DR EMBL; U51627; G1695803; -
DR EMBL; U60415; G1698576; -
DR MIM; 602550; -
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF0010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
KW REPEAT; DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION;
KW ALTERNATIVE SPLICING.
FT DNA_BIND 30 42 BASIC DOMAIN.
FT DOMAIN 43 83 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT REPEAT 103 170 PAS-1.
FT REPEAT 285 351 PAS-2.
FT DOMAIN 358 401 PAC MOTIF.
FT VARSPLIC 1 4 MINI -> MADQRMDSSTISDFMSPGPTDLSSSLGTSGV
FT VARSPLIC 1 16 DCMRRKGSSTDIQ (IN ISOFORM BMAL1B).
FT VARSPLIC 1 16 MINIESMDTDKDDPHG -> MSKEAVSLWALTVSLOPPVPL
FT VARSPLIC 1 16 CVCREMTEGSGRRKQCVTLPLFISRELCTYLLRPPPP (IN ISOFORM MOP3).
FT VARSPLIC 181 181 T -> R (IN ISOFORM BMAL1C).
FT VARSPLIC 182 583 MISSING (IN ISOFORM BMAL1C).
FT VARSPLIC 231 348 MISSING (IN ISOFORM BMAL1D).
FT VARSPLIC 235 258 SECTHSTGYLKSMPFTKGLDED -> AFCTHSTGYFGI
FT FTTRSRHIVL (IN ISOFORM BMAL1E).
FT VARSPLIC 259 583 MISSING (IN ISOFORM BMAL1E).
FT VARSPLIC 400 483 ANVEGDEPTFPOLTAHSPHSDMLPSGEGPKRTHPTVPG
FT IPGSTRAGAGKIGRMAEIMEIHRIRGSSPSSCGSSPLNT
FT TS -> SRVDTHLGOVERCTVLSRPNRSFLIAGMTEPTS
FT WKAGTQPSHSSQHPPTAATACCPLEKVAQGRPTPLFQGFQG
FT EPGLOEK (IN ISOFORM BMAL1F).
FT MISSING (IN ISOFORM BMAL1F).
FT R -> G (IN REF. 2).
FT K -> R (IN AB000812).
FT S -> P (IN AB000816).
FT K -> N (IN AB000815).
FT D -> N (IN AB000815).
FT S -> N (IN AB000814).
FT SP -> LR (IN REF. 2).
SQ SEQUENCE 583 AA; 64206 MW; 37642A96 CRC32;

alignment_scores:
Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 1
Percent Similarity: 50.000 Percent Identity: 23.684

alignment_block:
US-09-030-606-223/rev x BMAL_HUMAN ..

Align seg 1/1 to: BMAL_HUMAN from: 1 to: 583

292 CCCCACCAAGTTAAATATATATGTCATTCACAGGAATCAAAATCTTT 243
|||:::|||||:::|||||:::|||||
14 ProHisGlyArgLeuGluTyrThrGlnHisGlnGlyArgIleLysAsnAl 30
242 TAGAATAGCACACATCCCAACAAGTGAKGGAACACTACTAATTCCTTGA 193
||| |||||:::|||||
30 aArgGluAlaHisSerGln..... 36
192 CTCCTTTGGCAGCATTTACTTTTGATAGAGAAGTCTCCAAATTAATACAA 143
37:::|||||:::|||||:::|||||
142 AATTTGGCAGACATTTTAATCTTGTCAAGACAATGTAGGAATGCC 93
:::|||||:::|||||
47 SerPheIleAspGluLeuAlaSerLeuValProThrCysAsnAlaMetSe 63
92 CCAMAAYAYWATTAATTGACCATTAAT 65
::: |||||:::|||||
63 rArgLysLeuAspLysLeuThrValLeu 72

Ratio: 1.605 Gaps: 3
Percent Similarity: 53.521 Percent Identity: 25.352

alignment_block:

US-09-030-606-223/rev x RRPO_TMGMV ..

Align seg 1/1 to: RRPO_TMGMV from: 1 to: 1608

```
238 ATAGCAGACTCCAAACAAGTGAGGAGACACTACTAATTCCTTAGACTTC 189
   :: ||||| ||||| :: |||
1320 ValTyrHisSerLysGlnIleAsnGlyIleLeu.....AlaG1 1332

188 CTTTGGCAGCATTA..... 173
   ::|||::|||
1332 yPheSerGluLeuThrArgLeuLeuGluAlaPheAspSerLysLysP 1349

172 .....TTTGATAGAGAGTCTCCAAATAAATACAAATTTGGCACA 131
   ||| ::|||::|||::|||::|||
1349 heLeuPhePheThrArgLysThrProGluGlnIleGlnGluPheSer 1365

130 GACATTTTAATCTTGTCAAGACAATGTAAGAAATGCCCAAAAYATWAT 81
   |||:: ::|||::|||
1366 AspLeu.....AspSerHisValProMetAspValLe 1376

80 TAAATTGACCATA 68
   ::||| |||
1376 uGluLeuAspIle 1380
```

seq_name: SwissProt_37:YJ09_YEAST

seq_documentation_block:

ID YJ09_YEAST STANDARD; PRT; 1121 AA.
AC P47107;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 127.4 KD PROTEIN IN RAD26-GEF1 INTERGENIC REGION.
GN YJR039W OR J1614.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA HUANG M.-E., CHUAT J.-C., GALIBERT F.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49538; G1015689; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1121 AA; 127438 MW; F5B4818A CRC32;

alignment_scores:

Quality: 59.50 Length: 79
Ratio: 1.352 Gaps: 3
Percent Similarity: 55.696 Percent Identity: 27.848

alignment_block:

US-09-030-606-223/rev x YJ09_YEAST ..

Align seg 1/1 to: YJ09_YEAST from: 1 to: 1121

```
283 AAGTTAAATATATATGTTCATTCAGGAAATCAAATCTTTAGAAATAGC 234
   ::|||::|||::|||::|||::|||
885 GlnLeuAsnTyrSerValHisAspAsnLysPheSerIleGluGlnValSe 901
```

```
233 ACACCTCCAAACAAGTGAGGAGACACTACTAATTCCTTAGACTTCCTTG 184
   ::|||::|||::|||
901 rasnArgLeuAsnIleSerGlyLe.....ThrIleThrS 913

183 GCAGCATTACTTTGTATAG.....AAGTCACCAATAAATACAA 143
   ::||| ||||| ||| ::||| |||
913 erSerIlePhePheAspLysArgLysAlaLysMetAlaArgLysGlnGln 929

142 AATTTGGCACA.....GACATTTTAATCTTGTCAAGACA 108
   |||::|||::|||
930 AsnIleGlyThrTrpValTyrLeuGluGluMetIleLeuAspValAr 946

107 ATGTAAGAAATGCCCAAAAYATWATTAATGACC 71
   : ||| ||| ||| |||
946 glyGlyValMetArgPheAsnValIleHisThrThr 958
```

seq_name: SwissProt_37:RFAK_ECOLI

seq_documentation_block:

ID RFAK_ECOLI STANDARD; PRT; 357 AA.
AC P27242;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LIPOPOLYSACCHARIDE 1,2-N-ACETYLGALACTOSAMINE TRANSFERASE (EC 2.4.1.56).
GN RFAK.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 92325067.
RA KLENA J.D., PRADEL E., SCHNAITMAN C.A.;
RT "Comparison of lipopolysaccharide biosynthesis genes rfaK, rfaL,
RT rfaY, and rfaZ of Escherichia coli K-12 and Salmonella typhimurium."
RL J. BACTERIOL. 174:4746-4752(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 94316500.
RA SOFIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL NUCLEIC ACIDS RES. 22:2576-2586(1994).
CC -1- FUNCTION: ADDS A N-ACETYL-D-GLUCOSAMINE GROUP ON A GLUCOSE
CC GROUP OF LPS.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE +
CC LIPOPOLYSACCHARIDE -> UDP + N-ACETYL-D-GLUCOSAMINYL-
CC LIPOPOLYSACCHARIDE.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE.
CC -1- SIMILARITY: SHOWS VERY LITTLE SIMILARITY TO S.TYPHIMURIUM RFAK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95398; G147573; -
DR EMBL; U00039; G466761; -
DR EMBL; AE000440; G1790053; -
DR PIR; C42981; C42981.
DR ECOGENE; EG11423; RFAK.
KW LIPOPOLYSACCHARIDE BIOSYNTHESIS; GLYCOSYLTRANSFERASE; TRANSFERASE.
FT CONFLICT 293 293 I -> M (IN REF. 1).
SQ SEQUENCE 357 AA; 41729 MW; 7DF52EFD CRC32;

alignment_scores:

Quality: 58.50 Length: 96
Ratio: 1.147 Gaps: 3
Percent Similarity: 53.125 Percent Identity: 23.958

alignment_block:

US-09-030-606-223/rev x RFAK_ECOLI ..

Align seg 1/1 to: RFAK_ECOLI from: 1 to: 357

```

283 AGGTTAAATATATGTCATTC.....AGGAATCAAAATCTTT 243
|||::||| ||| |||:::|
16 LysileLysileasnPheluserPhelupheArGAsnLysileAsnAs 32
242 TAGAATAGCACACTCCAAACAGTAGAGGACACTACTAATTCCTTGA 193
:::|:::| ||||| ::|:::|
32 nGlnleThrasProAlaGlnValLysSerCysleullele..... 46
192 CTTCCTTTGGCAGCACTTACTTTTGATAAGAGTCTCCAATAAAATCAA 143
47 .....HisAspAsnAsnLysLeu... 52
142 AATTTGGCACAGACATTTTAATCTGTCAAGACAATGTAAGAAATGCC 93
|||:::|:::|:::|:::|:::|
53 .....GlyAspleulleValLeuserSerileTyrArgGluLeuTy 66
92 CCAMAAYAWATTAATGACCATATATACCAATATCAGTCCCTAAGATA 43
:::|:::|:::|:::|:::|:::|
66 rSerLysGlyValLysileThrLeuThrAsnArgLysGlyGlyGluP 83
42 ATTTTCTGAATGAGAATGTTTGTGTTGTTGTTG 5
:::|:::|:::|:::|:::|
83 heLeuserAsnAsnLysAsnlePheGluPhecysile 95

```

seq_name: SwissProt_37:SECY_SYNP7

seq_documentation_block:

```

ID SECY_SYNP7 STANDARD; PRT; 439 AA.
AC P31159;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PREPROTEIN TRANSLOCASE SECY SUBUNIT.
GN SECY.
OS SYNECHOCOCUS SP. (STRAIN PCC 7942) (ANACYSTIS NIDULANS R2), AND
OS SYNECHOCOCUS SP. (STRAIN PCC 6301).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 7942;
RX MEDLINE; 93041999.
RA NAKAI M., TANAKA A., OMATA T., ENDO T.;
RT "Cloning and characterization of the secy gene from the
RL cyanobacterium Synechococcus PCC7942."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 6301;
RX MEDLINE; 97444291.
RA SUGITA M., SUGISHITA H., FUJISHIRO T., TSUBOI M., SUGITA C., ENDO T.,
RT "Organization of a large gene cluster encoding ribosomal proteins in
RT the cyanobacterium Synechococcus sp. strain PCC 6301: comparison of
RT gene clusters among cyanobacteria, eubacteria and chloroplast
RL genomes."
RN GENE 195:73-79(1997).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
CC TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
CC BY FORMING PART OF A CHANNEL.
CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).

```

DR EMBL; X68056; G47618; -
DR EMBL; AB000111; D1023332; -
DR PIR; S27155; S27155.
DR PROSITE; PS00755; SECY_1; 1.
DR PROSITE; PS00756; SECY_2; 1.
DR PRAM; PF00344; secy; 1.
KW PROTEIN TRANSPORT; TRANSLOCATION; TRANSMEMBRANE.
SQ SEQUENCE 439 AA; 47150 MW; 6193F057 CRC32;

```

alignment_scores:

Quality: 58.50 Length: 76
Ratio: 1.773 Gaps: 1
Percent Similarity: 43.421 Percent Identity: 25.000

alignment_block:

US-09-030-606-223 x SECY_SYNP7 ..

Align seg 1/1 to: SECY_SYNP7 from: 1 to: 439

```

184 CAAGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGAGTGT 233
:::|:::|:::|:::|:::|:::|
121 ArgArgLysileAlaGlnLeuThrArgTyrValSerLeuGlyTrpAlaLe 137
234 GCTATTCTAAGATTGATTTCCTGGAATGACAATTAATTTAACT. 282
::|:::|:::|:::|:::|
137 uLeuGlnSerileValIleAlaValTrpValThrArgTyrAlaValThrP 154
282 .....
154 roGlyProLeuPheThrIleGlnThrAlaLeuAlaLeuValAlaGlySer 170
283 .....TTGGTGGGGGAANAGTTATAGGACC 308
|||:::|:::|
171 MetPheValMetTrpIleSerGluLeuIleThrGluArgGlyIleGlyAs 187
309 ACAGCTCTCATTCTGATCTGTAAT 336
:::|:::|:::|:::|:::|
187 nGlyAlaSerLeuLeuIlePheLeuAsn 196

```

seq_name: SwissProt_37:PEP3_YEAST

seq_documentation_block:

```

ID PEP3_YEAST STANDARD; PRT; 918 AA.
AC P27801;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE VACUOLAR MEMBRANE PROTEIN PEP3.
GN PEP3 OR VPS18 OR YLR148W OR U9634.2.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92049305.
RA PRESTON R., MANOLSON M.F., BECHERER K., WEINDHAMMER E.,
RA KIRKPATRICK D., WRIGHT R., JONES E.W.;
RT "Isolation and characterization of PEP3, a gene required for vacuolar
RT biogenesis in Saccharomyces cerevisiae."
RL MOL. CELL. BIOL. 11:5801-5812(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92049306.
RA ROBINSON J.S., GRAHAM T.R., EMR S.D.;

```



```

206 TAGTGTCCMTCACCTGTTG.....GAGT 231
    ::::: |||||
75 rargmetHrSerProcysLeuLeuValArgAspLeuGluValIleIySH 92
232 GTGCTATCTAAAGATTGAT...TTCTGGAATGACAAATTATTTT 278
    ::::: |||||
92 IsIleMetIleLysAspPheGluValPheSerAspArgGlyLeuGluPhe 108
279 AACCTTGTGGG.....GGAANAAGTTA 301
    ::: |||
109 SerLysGluGlyLeuGlyGlnAsnLeuPheHisAlaAspGlyAspThrTr 125
302 TAGG.....ACCACAGTCTTCACCTTCTGATACCTTGTGA 333
    :|||
125 parGTThrLeuArgAsnArgPheThrProIlePheThrSerGlyLysLeuL 142
334 AATTAATCTTTAT 347
    || ::|||
142 ysAsnMetPheTyr 146

```

seq_name: SwissProt_37:TEGU_EBV

seq_documentation_block:

```

ID TEGU_EBV STANDARD; PRT; 3149 AA.
AC P03186;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE LARGE TEGUMENT PROTEIN.
GN BPLF1.
OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84270667.
RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HATFUL G., HUDSON G.S., SATCHELL S.C., SEGUIN C.,
RA TUFENELL P.S., BARRELL B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL NATURE 310:207-211(1984).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01555; G59094; .
CC PIR; A03747; Q0BE8.
CC PIR; S32993; S32993.
SQ SEQUENCE 3149 AA; 337954 MW; 8C0A19B4 CRC32;

```

alignment_scores:

```

Quality: 58.00 Length: 54
Ratio: 1.706 Gaps: 2
Percent Similarity: 62.963 Percent Identity: 31.481

```

alignment_block:

US-09-030-606-223 x TEGU_EBV ..

Align seg 1/1 to: TEGU_EBV from: 1 to: 3149

```

204 AGTAGTGTCCMTCACCTGTTTGAGAGTGCTATCTAAAGATTGTA 253
    ||:::|||||
682 SerAspIleProThrThr.....GluAspGluAs 691

```

```

254 TTTCGTGGAATGACATATATTTTAAAC...TTGTGGGGGGAANAAGTT 300
    ||:::|||||
691 pmetPheGluAspGluValPheSerAsnSerLeuGluSerGlySerSera 708
301 ATAGGACACAGTCTTCACCTTCTGATACCTGTAAATTAATCTTTATTGC 350
    ||::: |||||
708 laproThrSerProIleThrLeuAspThrAlaArgSerGlnTyrGln 724
351 ACTGTGTTTGAC 362
    ||| |||||
725 ThrThrPheAsp 728

```

seq_name: SwissProt_37:UNG_BORBU

seq_documentation_block:

```

ID UNG_BORBU STANDARD; PRT; 223 AA.
AC 051082; Q44841;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.-) (UDG).
GN UNG OR UDG OR BB0053.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 35210 / B31;
RX MEDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERIAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL NATURE 390:580-586(1997).
RN [2]
RP SEQUENCE OF 1-80 FROM N.A.
RX STRAIN-ATCC 35210 / B31;
RA GEBBIA J.A., BACKENSON P.B., ANDA P., COLEMAN J.L., BENACH J.L.;
RT SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001118; G2687922; .
CC EMBL; U57684; G1373351; -.
DR TIGR; BB0053; .
DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
DR HSSP; P13051; 1AKZ.
KW DNA REPAIR; HYDROLASE; GLYCOSIDASE.
FT ACT_SITE 67 67 GENERAL BASE (BY SIMILARITY).
FT CONFLICT 32 32 E -> D (IN REF. 2).
FT CONFLICT 73 79 NOANGLA -> KPSXWTC (IN REF. 2).
SQ SEQUENCE 223 AA; 25561 MW; EA56D0AE CRC32;

```

alignment_scores:

```

Quality: 57.50 Length: 97
Ratio: 1.127 Gaps: 4

```


RX MEDLINE; 89384829.
 RA AMSELEM S., DUQUESNOY P., ATTREE O., NOVELLI G., BOUSNINA S.,
 RA POSTELVINAY M.-C., GOOSSENS M.;
 RT "Laron dwarfism and mutations of the growth hormone-receptor gene.";
 RL NEW ENGL. J. MED. 321:989-995(1989).
 RN [5]
 RP VARIANTS LARON DWARFISM.
 RX MEDLINE; 93278381.
 RA AMSELEM S., DUQUESNOY P., DURIEZ B., DASTOT F., SORBIER M.-L.,
 RA VALLEIX S., GOOSSENS M.;
 RT "Spectrum of growth hormone receptor mutations and associated
 RT haplotypes in Laron syndrome.";
 RL HUM. MOL. GENET. 2:355-359(1993).
 RN [6]
 RP VARIANT LARON DWARFISM HIS-170.
 RX MEDLINE; 94185645.
 RA DUSQUESNOY P., SORBIER M.-L., DURIEZ B., DASTOT F., BUCHANAN C.R.,
 RA SAVAGE M.O., PRECE M.A., CRAESCU C.T., BLOUQUIT Y., GOOSSENS M.,
 RA AMSELEM S.;
 RT "A single amino acid substitution in the exoplasmic domain of the
 RT human growth hormone (GH) receptor confers familial GH resistance
 RT (Laron syndrome) with positive GH-binding activity by abolishing
 RT receptor homodimerization.";
 RL EMBO J. 13:1386-1395(1994).
 RN [7]
 RP VARIANTS IDIOPATHIC SHORT STATURE LYS-62; CYS-179 AND ASP-242.
 RX MEDLINE; 96013502.
 RA GODDARD A.D., COVELLO R., LUOH S.-M., CLACKSON T., ATTIE R.M.,
 RA GESUNDHEIT N., RUNDLE A.C., WELLS J.A., CARLSSON L.M.S.;
 RT "Mutations of the growth hormone receptor in children with idiopathic
 RT short stature. The Growth Hormone Insensitivity Study Group.";
 RL NEW ENGL. J. MED. 333:1093-1098(1995).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 19-256.
 RX MEDLINE; 92196577.
 RA DE VOS A.M., ULTSCH M., KOSSIAKOFF A.A.;
 RT "Human growth hormone and extracellular domain of its receptor:
 RT crystal structure of the complex.";
 RL SCIENCE 255:306-312(1992).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 19-256.
 RX MEDLINE; 97113023.
 RA SUNDSTROM M., LUNDQVIST T., RODIN J., GIEBEL L.B., MILLIGAN D.,
 RA NORSTEDT G.;
 RT "Crystal structure of an antagonist mutant of human growth hormone,
 RT G120R, in complex with its receptor at 2.9-A resolution.";
 RL J. BIOL. CHEM. 271:32197-32203(1996).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR PITUITARY GLAND GROWTH HORMONE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DISEASE: DEFICIENCY IN GHR IS THE CAUSE OF PITUITARY DWARFISM II
 CC (LARON-TYPE PITUITARY DWARFISM OR LARON SYNDROME (LS)). IT ALSO
 CC CAUSES IDIOPATHIC SHORT STATURE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06562; G31738; -;
 DR EMBL; M28466; G183171; -;
 DR EMBL; M28458; G183171; JOINED.
 DR EMBL; M28459; G183171; JOINED.
 DR EMBL; M28460; G183171; JOINED.
 DR EMBL; M28461; G183171; JOINED.
 DR EMBL; M28462; G183171; JOINED.
 DR EMBL; M28463; G183171; JOINED.
 DR EMBL; M28464; G183171; JOINED.

DR EMBL; M28465; G183171; JOINED.
 DR PIR; S04530; S04530.
 DR PIR; A33991; A33991.
 DR PDB; 3HHR; 30-APR-94.
 DR PDB; 1HWG; 19-NOV-97.
 DR PDB; 1HWH; 19-NOV-97.
 DR PDB; 1AXI; 28-JAN-98.
 DR PDB; 1A22; 29-APR-98.
 DR MIM; 600946; -;
 DR MIM; 262500; -;
 DR PROSITE; PS00241; RECEPTOR_CYTOKINES_1; 1.
 DR PROSITE; PS00340; RECEPTOR_CYTOKINES_2; FALSE_NEG.
 DR PFAM; PF00041; fn3; 1.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; 3D-STRUCTURE;
 KW DWARFISM; DISEASE MUTATION.
 FT SIGNAL 1 18
 FT CHAIN 19 638 POTENTIAL.
 FT DOMAIN 19 264 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 265 288 POTENTIAL.
 FT DOMAIN 289 638 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 145 252 FIBRONECTIN TYPE-III.
 FT DISULFID 56 66
 FT DISULFID 101 112
 FT DISULFID 126 140
 FT CARBOHYD 46 46 POTENTIAL.
 FT CARBOHYD 115 115 POTENTIAL.
 FT CARBOHYD 156 156 POTENTIAL.
 FT CARBOHYD 161 161 POTENTIAL.
 FT CARBOHYD 200 200 POTENTIAL.
 FT VARIANT 62 62 E -> K (IN IDIOPATHIC SHORT STATURE).
 FT VARIANT 89 89 R -> K (IN LARON-DWARFISM).
 FT VARIANT 114 114 F -> S (IN LARON-DWARFISM).
 FT VARIANT 143 143 V -> A (IN LARON-DWARFISM).
 FT VARIANT 162 162 V -> D (IN LARON-DWARFISM).
 FT VARIANT 170 170 D -> H (IN LARON-DWARFISM; ABOLISH
 FT RECEPTOR HOMODIMERIZATION).
 FT VARIANT 179 179 R -> C (IN LARON-DWARFISM, AND IDIOPATHIC
 FT SHORT STATURE).
 FT VARIANT 229 229 R -> G (IN LARON-DWARFISM).
 FT VARIANT 242 242 E -> D (IN IDIOPATHIC SHORT STATURE).
 FT CONFLICT 544 544 I -> L (IN REF. 2).
 FT STRAND 53 58
 FT STRAND 64 68
 FT STRAND 82 88
 FT STRAND 99 100
 FT TURN 104 107
 FT TURN 109 110
 FT STRAND 111 114
 FT TURN 116 117
 FT STRAND 124 131
 FT TURN 132 133
 FT STRAND 134 142
 FT HELIX 143 145
 FT STRAND 147 147
 FT STRAND 153 162
 FT TURN 164 165
 FT STRAND 168 176
 FT TURN 179 180
 FT TURN 183 186
 FT STRAND 190 198
 FT TURN 199 200
 FT STRAND 205 206
 FT STRAND 210 210
 FT STRAND 221 221
 FT TURN 222 223
 FT STRAND 225 234
 FT STRAND 247 250
 SQ SEQUENCE 638 AA; 71499 MW; 5E92540C CRC32;

alignment_scores:

 Quality: 57.50
 Ratio: 1.983

 Length: 53
 Gaps: 2

Percent Similarity: 54.717 Percent Identity: 30.189

alignment_block:

US-09-030-606-223 x GHR_HUMAN ..

Align seg 1/1 to: GHR_HUMAN from: 1 to: 638

```
11 AACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGAGCTGATAT 60
   ::|||:::|||||
232 SerLysGlnArgAsnSer..... 237

61 TGGTAATTATGTCATTTAATWRTRTTKTGGGCATTTCTTACATGT 110
   ||| ||| ||| ||| ::|::| ||| |||
238 .GlyAsnTyrGlyGluPheSerGluValLeu.....TyrValThrLeup 252

111 CTGACACAGATTAAATGTCTGTGCCAAATTTGTATTTATTGAGA 160
   ::|::: ::||| ::|::| ||| |||
252 roGlnMetSerGlnPheThrCysGluGluAspPheTyrPheProTrpLeu 268

161 CTTCCTTATC 169
   |||::| |||
269 LeuIleIle 271
```


This Page Blank (uspto)

DR PFAM: PF00083; sugar_tr; 1.
SQ SEQUENCE 586 AA; 64272 MW; 3E637F95 CRC32;

alignment_scores:
Quality: 68.00 Length: 92
Ratio: 1.581 Gaps: 5
Percent Similarity: 46.739 Percent Identity: 26.087

alignment_block:
US-09-030-606-223 x Q07824 ..

Align seg 1/1 to: Q07824 from: 1 to: 586

```
91 GGGGCATTCTTACATGCTTGACAAAGATTAAATGCTGTGCCAAA 140
||| ::||| ||||| ::|||
414 GlyGluLeuProTyrIleAlaLeuIleIleGlyMetMetValCysAlaI 430
141 TTTTGATTTTATTTGGAGACTTCTTATCAAAAGTAAATGCTGCCAAGA 190
||| ::||| ::||| ::|||
430 aPheIleTrrPTyrMetaspAsnspTyrLeuLysArgCys.....ArgA 445
191 AGTCTAAGGAATTAGTA..... 207
::: ::|||
445 IalysGlyLysLeuValProGluAlaArgLeuTyrAlaMetValIleAla 461
208 .....GTGTCCCM..TCACTTGTTGGAGTGTGCTATTCTAAAGAT 248
||| ||||| ::|||
462 GlyThrValPheProIleGlyIleLeuTrrP..... 471
249 TTTGATTTCTCGAATGACATTAAT.....ATTTTAA 280
::: ||| ::|||
472 ...PheCysTrpThrGlyTyrTyrProHisLysIleHisTrpMetValP 487
281 CTTTGGTGGGGGAANAGTTATAGGA 306
||| ||||| ::|||
487 rorThrValGlyGlyAlaPheIleGly 495
```

seq_name: sp_virus:039250

seq_documentation_block:

ID 039250 PRELIMINARY; PRT; 1080 AA.

AC 039250;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE COUNTERPART OF HSV-1 GENE UL52 AND VZV GENE 6.
GN 7.

OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.

RN [1]
RP SEQUENCE OF 1-198 FROM N.A.

RC STRAIN-NS80567;
RX MEDLINE; 94058670.

RA RIGGIO M.P., ONIONS D.E.;

RT "DNA sequence of a gene cluster in the equine herpesvirus-4 genome
RT which contains a newly identified herpesvirus gene encoding a
RT membrane protein."
RL Arch. Virol. 133:171-178(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NS80567;
RX MEDLINE; 98264497.

RA TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;

RT "The DNA sequence of equine herpesvirus-4."
RL J. Gen. Virol. 79:1197-1203(1998).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-NS80567;
RX MEDLINE; 98264497.

RA TELFORD E.A., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF030027; AAC59589.1; -.

SQ SEQUENCE 1080 AA; 120603 MW; 2A8793CB CRC32;

alignment_scores:
Quality: 66.00 Length: 125
Ratio: 1.179 Gaps: 8
Percent Similarity: 44.800 Percent Identity: 26.400

alignment_block:
US-09-030-606-223 x 039250 ..

Align seg 1/1 to: 039250 from: 1 to: 1080

```
3 AACAAACAACAAAAAACATCTTCATTCAGAAAAATATCTTAGGG 52
||||| ||| ||||| ||| ::|
773 AsnLysThrThrAspLysThrProLeuGln..... 782
53 ACTGATATTGGTAATTATGGTCAATTTAATWRRTTKTGGGGCATTTCT 102
||| ::||| ::|||
783 .....LeuLeuLeuGluThrAsn.....AlaC 790
103 TACATGTCTTGACAAAG.....ATTAAATGCTGTGCCAAA 140
::||| ::||| ||| ::|||
790 YsGlnCysThrGluLysMetGlyPheArgIleThrValProValProPro 806
141 TTTTGATTTTATTTGGAGACTTCTTATCAAAAGTAAATGCTGCCAAGA 190
||| ||||| ||| |||||
807 ProTyrIleLeuAlaGly.....ProGluAl 815
191 AGTCTAAGGAATTAGTAGTGT.....CCCMTCACCTGTTTGAG.... 230
::: ||| ::||| ::|||
815 aLeuArgGlyValAlaArgIleIleGlnGlnAlaValValLeuGluArgT 832
231 .....TGTCCTATCTAAAGATTTTGATTTCTCTGGAA 263
||| ::||| ::|||
832 hrPheThrGluSerMetCysSerValLeuArgAspPheSerPheLeuAsp 848
264 TGACAAATATATTTTAACCTTGTGGGGGAANAGTTATAGACACAGT 313
::||| ::||| ||| |||||
849 ThrGlyValTyrSerHis.....GlyArgSerLeuArgLeuProPh 862
314 CTTCACTTCTGATACCTGTAATTA 338
||| ||||| ::|||
862 ephe.....CysLysVal 866
```

seq_name: sp_bacteria:Q56394

seq_documentation_block:

ID Q56394 PRELIMINARY; PRT; 301 AA.

AC Q56394;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE XYNB.

GN XYNB.

OS Thermoanaerobacter thermosulfurogenes (Clostridium
OS thermosulfurogenes).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter
OC group; Thermoanaerobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-EM1 (DSM 3896);

RA MATUSCHER M., SAHM K., BAHL H.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; U50952; AAB08045.1; -.

DR PFAM; PF00528; BPD_transp; 1.

DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; 1.

KW Transport; Transmembrane.

SQ SEQUENCE 301 AA; 33755 MW; 6630324E CRC32;

alignment_scores:
Quality: 66.00 Length: 108

Ratio: 1.200 Gaps: 4
Percent Similarity: 50.926 Percent Identity: 24.074

alignment_block:

US-09-030-606-223 x Q56394 ..

Align seg 1/1 to: Q56394 from: 1 to: 301

```
70 TGGTCATTTAATWRTTKTGGGGCATTCCTTACATGTCTTGACAAG 119
   ::|||::: ::|||::: |||||::: |||
121 TyrSerMetTyrPheThrSerGlyLeuIleProTyrTyrLeuIleAr 137
120 ATTAAGATGCTGTGCCAAATTTGTATTATTGAGACTTCTTATC 169
   | ||| ::|||::: |||
137 gaspLeuHisLeuMetAsnAspPheLeuValTyr..... 148
170 AAAAGTAATGCTGCCAAAGAGTCTAAGCAATTAGTAGTTCCTCA 219
   ::|||:::
149 .....IleLeuProGly 152
220 CTTGTT.....TGGAGTGTCTATTCATAAGATTGATTCTCGCAA 263
   ::|||::: |||||::: |||||:::
153 IleValSerAlaTyrPasnIleMet.....ValValArgSerTyrI 166
264 TGACAAATTATTTAACTTTGGTGGGGGAANAGTTATAGCACACAGT 313
   |||::: ::|||::: |||
166 easpGlyLeuProAlaSerLeuValGluSerAlaLysIleaspGlyAla 183
314 CTCACATTCGTACTCTGTAATTAATCTTT.....ATTGCACTT 354
   |||::: |||||::: |||
183 ergIleuArgIleIlePheSerIleIlePheProLeuSerValProVal 199
355 GTTTTGACCACTTAAGCTATATGTT 378
   ::|||::: |||||::: |||
200 LeuAlaThrIleThrLeuPheVal 207
```

seq_name: sp_virus:Q90699

seq_documentation_block:

ID Q90699 PRELIMINARY; PRT; 1818 AA.
AC Q90699;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE HYPOTHETICAL 207.4 KD PROTEIN.
OS Beet virus Q.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98378056.
RA KOENIG R., PLEIJ C., BEIER C., COMMANDEUR U.;
RT "Genome properties of beet virus Q, a new furo-like virus from
RT sugarbeet, determined from unpurified virus.";
RL J. Gen. Virol. 79:2027-2036(1998).
DR EMBL; AJ223596; CAA11457.1;
DR PFM; PF00978; RNA_dep_RNAp012; 1.
KW Hypothetical protein.
SQ SEQUENCE 1818 AA; 207431 MW; 7003FEEF CRC32;

alignment_scores:

Quality: 65.50 Length: 95
Ratio: 1.394 Gaps: 5
Percent Similarity: 49.474 Percent Identity: 25.263

alignment_block:

US-09-030-606-223 x Q90699 ..

Align seg 1/1 to: Q90699 from: 1 to: 1818

```
90 TGGGGCATTTCCCTTACATGTCTTGACAAGATTAATGCTGTGCCAA 139
   |||::: ::|||::: |||||::: |||
1647 TrpSerAlaAlaLeuSerLeuIleGluSerLeuProLeuGluLysAlaAr 1663
```

```
140 ATTTGTATTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAG 189
   :||| |||||::: |||||::: |||
1663 gPheMetIlePheGlyGlyAspAsp...SerLeuIlePhePheProLysA 1679
190 AAGCTAAGCAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATT 239
   ::|||::: ::|||::: |||
1679 rgmet.....AlaIleProAspProCysArgArgLeuAlaAla 1691
240 CTAAGATTTTGCATTCCTGGAATGACAATTATATTAACTTTGCTGG 289
   ::|||::: |||||::: |||
1692 ValTrpAsnPheAspCys.....LysMetPheAsnPheGluAs 1704
290 G.....CGAANAGTTATAGCA 306
   : |||||
1704 nasAlaPheCysGlyLysPheLeuIleLysValGlyGluArgTyrArg. 1720
307 CCACAGCTTCACCTTCGATACCTGTAATTAATC 341
   |||::: |||||::: |||
1721 .....PheAlaProAspProTyrLysLeuLeu 1729
```

seq_name: sp_virus:Q56790

seq_documentation_block:

ID Q56790 PRELIMINARY; PRT; 1778 AA.
AC Q56790;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE 204 KDA PROTEIN.
OS beet soil-borne virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-AHLDM.
RX MEDLINE; 98062137.
RA KOENIG R., LOSS S.;
RT "Beet soil-borne virus RNA 1: genetic analysis enabled by a starting
RT sequence generated with primers to highly conserved helicase-encoding
RT domains.";
RL J. Gen. Virol. 78:3161-3165(1997).
DR EMBL; Z97873; CAB10764.1;
DR PFM; PF00978; RNA_dep_RNAp012; 1.
DR PFM; PF00149; STphosphatase; 1.
SQ SEQUENCE 1778 AA; 203690 MW; BE65D96F CRC32;

alignment_scores:

Quality: 64.50 Length: 95
Ratio: 1.372 Gaps: 5
Percent Similarity: 49.474 Percent Identity: 27.368

alignment_block:

US-09-030-606-223 x Q56790 ..

Align seg 1/1 to: Q56790 from: 1 to: 1778

```
90 TGGGGCATTTCCCTTACATGTCTTGACAAGATTAATGCTGTGCCAA 139
   |||::: ::|||::: |||||::: |||
1607 TrpSerAlaAlaLeuAlaLeuAspAlaLeuProLeuGluArgAlaLy 1623
140 ATTTGTATTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAG 189
   |||||::: |||||::: |||
1623 sPheMetValPheGlyGlyAspAsp...SerLeuValPhePheProLysA 1639
190 AAGCTAAGCAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATT 239
   ::|||::: |||||::: |||
1639 smet.....AsnLeuAlaAspProCysGlyArgLeuAlaSer 1651
240 CTAAGATTTTGCATTCCTGGAATGACAATTATATTAACTTTGCTGG 289
   |||::: |||||::: |||
1652 LeuTrpAsnPheAspCys.....LysPhePheAsnPheGluAs 1664
```



```

290 G.....GGAANAGTTATAGGA 306
      :
1664 nasMetPhecGlyGlyPheLeuLeuLysIleGlyGluAsnTyrLys. 1680
      :
307 CCACAGCTTCACCTCTGATACCTGTAATTAATC 341
      |||::: ||| |||||:::
1681 .....PheAlaProAspProPheLysLeu 1689

seq_name: sp_bacteria:084258

seq_documentation_block:
ID 084258 PRELIMINARY; PRT; 414 AA.
AC 084258;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 45.8 KD PROTEIN.
GN CT256.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL Science 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AE001298; AAC67849.1; -.
DR Hypothetical protein.
KW
SEQUENCE 414 AA; 45838 MW; 0B5C5C4D CRC32;

```

```

alignment_scores:
    Quality: 64.50
    Ratio: 1.075
    Percent Similarity: 57.692
    Percent Identity: 29.808
    Gaps: 6
alignment_block:
    US-09-030-606-223 x 084258 ..
Align seg 1/1 to: 084258 from: 1 to: 414

106 ATGTCTTGACACAGATTAAATGTCTGTGCCAAATTTTGA..... 147
    :::::||||| :: |||||:::::|||||
7 ValilleuThrAlallePheValleuCysserGlyPheValSerleuSe 23
148 .....TTTATTTCGAGACT.....TCTT 166
    ||| ||| ::| |||
23 rHisIleAlaLeuPheSerleuProSerSerleuIleAlaHisIleTySerH 40
167 ATCAAAAGTAATGCTGCCAAAGGAAGCTTAAGCAATTAGTAGTGTCCM 216
    :::::|||| |::::: ::|||:::::||||
40 lsserLyssAsnArgGlnleuValArgGlnIleAlaAsnleuMetAlaTyPro 56
217 TCA...CTGTTTGAGAGTGTCTATTCTAAAGATTTTGATTTCCTGGA 263
    ::| |::: :::::|||| | | |:::
57 AsnHisleuLeuMetThrleuValPhePheAspIleGlyIleAsnIleG1 73
264 T...GACAATTATATTTAACCTTTGTGGGGGAANAAGTTATAGACAC 310
    : ::||| || ||||||||||::: :
73 yValGlnAsnCysIleAlaThrleuValGlyasp.....S 85
311 AGCTCTCACTTCTGATACTGTAAATTATCTTTATTCGACTGTGTTG 360
    :::::|||||::: ||::: ::| |:::||||

```

```

85 erAlaserleuLeuThrvAlgly.....ValProleuAlaleu 98
361 ACCATTAGCTA 372
|||::: |||
99 ThrleuValleu 102

seq_name: sp_plant:004582

seq_documentation block:
ID 004582 PRELIMINARY; PRT; 559 AA.
AC 004582:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE SIMILARITY TO TRITICUM AESTIVUM SERPIN.
GN F19R23.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA OSBORNE B.I., VYSOTSKAIA V.S., TORIUMI M., YU G., OJI O., SHEN Y.K
RA ARAUJO R., AU M., BUEHLER E., CONWAY A.B., CONWAY A.R., DEWAR K.,
RA FENG J., KIM C., KURTZ D., LI Y., SHINN P., SUN H., DAVIS R.W.,
RA ECKER J.R., FEERSPIEL N.A., THEOLOGIS A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC000375; AAB60763.1; -.
DR MENDEL; 16823; ARATH;1514;mn16823.
DR PFAM; PF00079; serpin; 4.
DR PROSITE; PS00284; SERPIN; 2.
KW Serpin.
SQ SEQUENCE 559 AA; 62722 MW; B8CC961A CRC32;

```

```

alignment_scores:
  Quality: 63.50      Length: 81
  Ratio: 1.549      Gaps: 3
Percent Similarity: 50.617      Percent Identity: 28.395

alignment_block:
US-09-030-606-223 x 004582  ..

Align seg 1/1  to: 004582  from: 1  to: 559

123  AAATGCTGTGCCAAATTTGTATTATTGGA..... 158
||||:  ::||| |||||  |||  |||||
467  LysPheArgIleProLysPheLysIleGluPheGlyPheGluAlaSerSe 483
159  .....CACTTCCTATCAAAAGTAAATGC 180
      ::|||  ::|||:::
483  rAlaPheSerAspPheGluLeuAspValSerPheTyrGlnLysThrLeu 500
181  TGGCAAAAGGAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGT..... 224
      ::  ::  |||  ::::: |||  |||:::
500  legIuIleAspGluLysGlyThrGluAlaValThrPheThrAlaPheArg 516
225  .....TTGAGTGTGCTATTCTAAAGATTTTGATTTCCTGGAATG 265
      |||  |||||::: |||  ::||| |||::: |
517  SerAlaTyrLeuGlyCysAlaLeuValLysProIleAspPheValAla.A 533
266  ACAATTATATTTTAACCTTTGGTGGGGAANAAGTTATAGGA 306
      ||:::  ::|||  |||:::  |||  |||
533  sPhIspProPheLeuPheLeuIleArgGluGluGlnThrGly 546

seq_name: sp_bacteria:Q59319

seq_documentation_block:
ID  Q59319      PRELIMINARY;      PRT;      825 AA.
AC  Q59319;
DT  01-NOV-1996 (TREMBlrel. 01, Created)

```


DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE ALPHA-DEXTRIN 6-GLUCANOHYDROLASE (EC 3.2.1.41)
DE (ALPHA-DEXTRIN ENDO-1,6-ALPHA-GLUCOSIDASE) (PULULANASE)
DE (PULULAN 6-GLUCANOHYDROLASE) (LIMIT DEXTRINASE) (DEBRANCHING ENZYME)
DE (AMYLOPECTIN 6-GLUCANOHYDROLASE).
GN PULA.
OS Caldocellum saccharolyticum (Caldicellulosiraptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiraptor.
RN [1]
RP SEQUENCE FROM N.A.
RA ALBERTSON G.D., MCHALE R., GIBBS M.D., BERGQUIST P.L.;
RL Eur. J. Biochem. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: STARCH-DEBRANCHING ENZYME, HYDROLYSES
CC (1-6)-ALPHA-GLUCOSIDIC LINKAGES IN PULULAN AND STARCH TO FORM
CC MALTOSE.
DR EMBL; L39876; AAB06264.1; -.
DR PFAM; PF00128; alpha-amylase; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 825 AA; 95732 MW; E89A4E9E CRC32;

alignment_scores:
Quality: 63.00 Length: 49
Ratio: 2.032 Gaps: 3
Percent Similarity: 63.265 Percent Identity: 36.735

alignment_block:
US-09-030-606-223/rev x Q59319 ..

Align seg 1/1 to: Q59319 from: 1 to: 825

238 ATAGCACACTCCAAACAA.....GTGAKGGGACACTACTAATTCCTTA 195
::: ||| |||::: |||::: |||::: |||
374 LeuLeuHisLeuLysGluLeuGlyValThrHisValHisLeuLeuProI1 390
194 GACTTCCTTTGGCAGCATTTACTTTGATTAAGAAGTCTCCAAATAAATAC 145
::: ||| |||::: ||| |||::: ||| |||
390 eSerAspPheGlySerVal.....AspAspLysAsnProAspLysArgT 405
144 AAAATTTTGGCAGACATTTTAATCTGTCAAGACAATGTAAGAA 98
|||::: ||| |||::: ||| ||| ||| |||
405 yrasnTrpGlyTyrAspProValLeuTyr.....GlnCysProGlu 418

seq_name: sp_plant:Q9ZWB6

seq_documentation_block:

ID Q9ZWB6 PRELIMINARY; PRT; 534 AA.
AC Q9ZWB6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE F21M11.11 PROTEIN.
GN F21M11.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., CONN L., HANSEN N.F.,
RA ALTAFI H., ARAUJO R., HUIZAR L., ROWLEY D., BUEHLER E., DUNN P.,
RA GONZALEZ A., KREMENTSKAIA I., KIM C., LENZ C., LI J., LIU S.,
RA LUKOS S., SCHWARTZ J., SHINN P., TORIUMI M., VYSOTSKAIA V.S.,
RA WALKER M., YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC003027; AAD10674.1; -.
SQ SEQUENCE 534 AA; 61847 MW; D30231CF CRC32;

alignment_scores:
Quality: 63.00 Length: 72
Ratio: 1.800 Gaps: 4
Percent Similarity: 48.611 Percent Identity: 36.111

alignment_block:
US-09-030-606-223/rev x Q9ZWB6 ..

Align seg 1/1 to: Q9ZWB6 from: 1 to: 534

292 CCCCACCAAG...TTAAATATATTTGTCATTCAGGAATCAAAATC 246
||||| ||| ||||| ::||| ||| :::::
129 ProProAlaLysGluLeuLysGluGlnCysLeuSerMetValAspGlnVa 145
245 TTTAGATAGCACACTCCAAACAAGTGAKGGGACACTACTAATTCCTT 196
||||||| ||| ||| |||::: ||| |||
145 lpheArgile.....TyrLysAlaValHisPheThrIleLeuValProL 160
195 AGACTTCCTTTGGCAGCATTTACT..... 173
::: ||| ||| ||| |||
160 euAlaGluPheLysSerIleThrLysGlnValCysLysLeuProSerPhe 176
172TTGATTAAGAAGTCTCCAAATAAATACAAATTTT 138
||| ||| ::||| |||
177 LeuSerProAlaLeuPheArgLysIleAspProAsn..... 188
137 TGGCAGACATTTTA 122
|||||||:::
189 .CysThrAspIleVal 193

seq_name: sp_virus:Q85476

seq_documentation_block:

ID Q85476 PRELIMINARY; PRT; 285 AA.
AC Q85476;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE (RECOVERED INSERTION MUTANT OF B77) SRC (FRAGMENT).
OS Rous sarcoma virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83155663.
RA MARDON G., VARMS H.E.;
RT "Frameshift and intragenic suppressor mutations in a Rous sarcoma
RT provirus suggest src encodes two proteins.";
RL Cell 32:871-879(1983).
DR EMBL; K03381; AAA42570.1; -.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
FT NON_TER 285
SQ SEQUENCE 285 AA; 31625 MW; 0238A3B1 CRC32;

alignment_scores:
Quality: 63.00 Length: 30
Ratio: 2.739 Gaps: 0
Percent Similarity: 76.667 Percent Identity: 43.333

alignment_block:
US-09-030-606-223/rev x Q85476 ..

Align seg 1/1 to: Q85476 from: 1 to: 285

286 CCAAGTTAAATATATATTTGTCATTCAGGAATCAAAATCTTTAGAAAT 237
||| ||||| ::||| ||||| |||::: |||:::
175 ProGlyLeuLysArgThrCysProSerArgLysGluAsnAlaCysArgLe 191
236 AGCACACTCCAAACAAGTGAKGGGACACTACTAATTCCT 197
::: ::::: |||::: ||| ||| ||| |||
191 userThrThrArgLysValThrGlyGlyTrpLeuIlePro 204


```

57 PROHISGLYARGLEUGLUTYRALAGLUHISGLINGLYARGILELYSASNAI 73
242 TAGAATAGCACACTCCAAACAAGTGAKGGGAACTACTAATTCCTTAGA 193
73 aArgGluaIaHisSerGln..... 79
192 CTCCTTTGGCAGCATTTCTTGATAGAGGCTCCAATAATAATACAA 143
80 .....ileglulysArgArgAspLysMetasn 89
142 AATTTGGCACAGACATTTTAATCTTGTCAGACATGTAAGGAATGCC 93
90 SerPheileaspgluleuAlaserleuValProthnCysasnaIametse 106
92 CCAMAAYAYWATTAATGACCATAAT 65
106 rArgLysLeuAspLysLeuThrValleu 115
```

```

92 CCAMAAYAYWATTAATGACCATAAT 65
106 rArgLysLeuAspLysLeuThrValleu 115
```

seq_name: sp_rodent:088810

seq_documentation_block:

| | | | | |
|----|--------|--------------|------|---------|
| ID | 088810 | PRELIMINARY; | PRT; | 626 AA. |
|----|--------|--------------|------|---------|

AC 088810;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)

DE BMAL1B.

GN BMAL1B.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RX MEDLINE; 98407895.

RA HONMA S., IKEDA M., ABE H., TANAHASHI Y., NAMIHIRA M., HONMA K., NOMURA M.;

RT "Circadian oscillation of BMAL1, a partner of a mammalian clock gene

RT Clock, in rat suprachiasmatic nucleus.";

RL Biochem. Biophys. Res. Commun. 250:83-87(1998).

DR EMBL; AB012600; BAA33450.1; -.

DR PFAM; PF00010; HLH; 1.

DR PFAM; PF00989; PAS; 1.

SQ SEQUENCE 626 AA; 68531 MW; 60BF0E7D CRC32;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 62.50 | Length: | 76 |
| Ratio: | 1.689 | Gaps: | 1 |
| Percent Similarity: | 48.684 | Percent Identity: | 23.684 |

alignment_block:

US-09-030-606-223/rev x 088810 ..

Align seg 1/1 to: 088810 from: 1 to: 626

```

292 CCCCCACCAAGTTAAATATATATTGTCATTCAGGAATCAAAATCTTT 243
57 PROHISGLYARGLEUGLUTYRALAGLUHISGLINGLYARGILELYSASNAI 73
242 TAGAATAGCACACTCCAAACAAGTGAKGGGAACTACTAATTCCTTAGA 193
73 aArgGluaIaHisSerGln..... 79
192 CTCCTTTGGCAGCATTTCTTGATAGAGGCTCCAATAATAATACAA 143
80 .....ileglulysArgArgAspLysMetasn 89
142 AATTTGGCACAGACATTTTAATCTTGTCAGACATGTAAGGAATGCC 93
90 SerPheileaspgluleuAlaserleuValProthnCysasnaIametse 106
```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:50 ; Search time 1809.22 Seconds
(without alignments)
417.572 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383

Sequence: 1 AAAACAACAACAACAAAAA.....ATTACGTATATGTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 376.8 | 98.4 | 400 | 36 | AA653016 | AA653016 ns68g01.s |
| 2 | 374 | 97.7 | 387 | 35 | AA579247 | AA579247 nf36d10.s |
| 3 | 369.8 | 96.6 | 522 | 26 | W56075 | W56075 zc56b05.r1 |
| 4 | 347.2 | 90.7 | 469 | 43 | AI188401 | AI188401 qd08e02.x |
| 5 | 339.4 | 88.6 | 432 | 35 | C75566 | C75566 C75566 Huma |
| 6 | 334.4 | 87.3 | 472 | 41 | AI027536 | AI027536 ow52c09.x |
| 7 | 327.4 | 85.5 | 389 | 39 | AA879150 | AA879150 nw84e05.s |
| 8 | 320.6 | 83.7 | 442 | 43 | AI160743 | AI160743 qb49d03.x |
| 9 | 317.4 | 82.9 | 444 | 47 | AI522136 | AI522136 tl78f03.x |
| 10 | 313.4 | 81.8 | 419 | 35 | C75584 | C75584 C75584 Huma |
| 11 | 303.4 | 79.2 | 496 | 41 | AI005274 | AI005274 ou07f12.x |
| 12 | 291.6 | 76.1 | 473 | 41 | AI051146 | AI051146 oy49d04.x |
| 13 | 288 | 75.2 | 407 | 36 | AA622892 | AA622892 np58c12.s |
| 14 | 280.4 | 73.2 | 344 | 28 | C14656 | C14656 C14656 Clon |
| 15 | 277 | 72.3 | 463 | 46 | AI420653 | AI420653 tf13b04.x |
| 16 | 261.2 | 68.2 | 324 | 31 | AA318914 | AA318914 EST21101 |
| 17 | 257 | 67.1 | 442 | 45 | AI382141 | AI382141 te30c04.x |
| 18 | 229.6 | 59.9 | 231 | 36 | AA653045 | AA653045 ns71f07.s |
| 19 | 227 | 59.3 | 268 | 28 | AA093999 | AA093999 cl1544.se |
| 20 | 213 | 55.6 | 225 | 36 | AA652561 | AA652561 ns73a06.s |
| 21 | 203 | 53.0 | 332 | 36 | AA657851 | AA657851 nu08d07.s |
| 22 | 189.6 | 49.5 | 352 | 28 | AA093469 | AA093469 kk5000.se |
| 23 | 185.4 | 48.4 | 246 | 36 | AA641128 | AA641128 nr28h07.r |
| 24 | 185 | 48.3 | 372 | 43 | AI214972 | AI214972 qm31f09.x |
| 25 | 181.4 | 47.4 | 200 | 32 | AA372885 | AA372885 EST84852 |
| 26 | 171 | 44.6 | 373 | 35 | AA580097 | AA580097 nh51h04.s |
| 27 | 158.4 | 41.4 | 282 | 30 | AA216139 | AA216139 hp0520.se |
| 28 | 151.4 | 39.5 | 255 | 24 | H83872 | H83872 yv84b03.s1 |
| 29 | 130.8 | 34.2 | 410 | 21 | T61899 | T61899 yb96e01.s1 |
| 30 | 118.2 | 30.9 | 703 | 48 | AI546846 | AI546846 PN2.1.07_ |
| 31 | 102.8 | 26.8 | 110 | 34 | AA503672 | AA503672 ng78h10.s |
| 32 | 101.2 | 26.4 | 309 | 21 | T61960 | T61960 yb96e01.r1 |
| 33 | 92.4 | 24.1 | 576 | 39 | C85936 | C85936 C85936 Mous |
| 34 | 92.4 | 24.1 | 438 | 44 | AI313916 | AI313916 uj38g06.x |
| 35 | 92.4 | 24.1 | 373 | 46 | AI450559 | AI450559 mq79f08.x |
| 36 | 91 | 23.8 | 425 | 29 | AA137503 | AA137503 mq79f07.r |
| 37 | 83.2 | 21.7 | 231 | 36 | AA658025 | AA658025 nu15d01.s |
| 38 | 76 | 19.8 | 275 | 49 | AV003262 | AV003262 AV003262 |
| 39 | 68.6 | 17.9 | 210 | 49 | AV010724 | AV010724 AV010724 |
| 40 | 64 | 16.7 | 298 | 50 | AV044015 | AV044015 AV044015 |
| 41 | 62.4 | 16.3 | 194 | 45 | AI350423 | AI350423 qtl17d05.x |
| 42 | 42.6 | 11.1 | 427 | 42 | AI152967 | AI152967 ud54d06.r |
| 43 | 42.6 | 11.1 | 437 | 46 | AI428145 | AI428145 ml45c02.x |
| 44 | 41.2 | 10.8 | 412 | 44 | AI289102 | AI289102 qw21g01.x |
| 45 | 40.8 | 10.7 | 392 | 23 | D61957 | D61957 HUM226F10B |

ALIGNMENTS

RESULT 1
AA653016
LOCUS AA653016 400 bp mRNA
DEFINITION ns68g01.s1 NCI_CGAP_Pt2 Homo sapiens CDNA clone IMAGE:1188816, mRNA
ACCESSION AA653016
NID 92584668
VERSION AA653016.1 GI:2584668

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 400)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT On Sep 12, 1996 this sequence version replaced gi:1325236.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 678 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 376.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .400 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1188816"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMPI0; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMPI0 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

```

| | | | | |
|------------|-------|------|------|-------|
| BASE COUNT | 130 a | 55 c | 70 g | 145 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 98.4%; | Score 376.8; | DB 36; | Length 400; |
| Best Local Similarity | 97.9%; | Pred. No. 3.4e-68; | | |
| Matches 375; Conservative | 5; | Mismatches 3; | Indels 0; | Gaps 0; |

QY 1 AAACAACAACAACAAAAACAATTCCTCATTCAGAAAAATTATCTTAGGGACTGATAT 60
 |||
 Db 14 AAACAACAACAACAAAAACAATTCCTCATTCAGAAAAATTATCTTAGGGACTGATAT 73
 QY 61 TGGTAATTATGGTCAATTTAATWRTTKTGGGGCATTTCCCTACATTGCTTGACAAGA 120
 |||
 Db 74 TGGTAATTATGGTCAATTTAATAATATTTGGGGCATTTCCCTACATTGCTTGACAAGA 133
 QY 121 TTAAATGCTGTGCCAAAAATTTGTATTTTAATTGGAGACTTCTATCAAAAAGTAATGC 180
 |||
 Db 134 TTAAATGCTGTGCCAAAAATTTGTATTTTAATTGGAGACTTCTATCAAAAAGTAATGC 193
 QY 181 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTC 240
 |||
 Db 194 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCATCACCTGTTGGAGTGTGCTATTC 253
 QY 241 TAAAGATTGATTTCTCGAATGACAATATATTTTAACCTTTGGTGGGGGAAANAGTT 300

| DB | Sequence | Position |
|----|---|----------|
| Db | 254 TAAAGATTTGATTTCCCTGGAATGACAAATTATATTTTAACTTGGTGGGGGAAGACT | 313 |
| QY | 301 ATAGAGCCACAGCTCTCACTCTGATFACCTGTAATTAATCTTTTATGTGCACCTGTTTG | 360 |
| Db | 314 ATAGGACCACAGCTCTTCACTCTGTAICTTGTAATTAATCTTTTATTGCACTGTTTG | 373 |
| QY | 361 ACCATTAAAGCTATATGTTAAA | 383 |
| Db | 374 ACCATTAAAGCTATATGTTAGAAA | 396 |

| | |
|------------|---|
| RESULT | 2 |
| AA579247 | |
| LOCUS | AA579247 |
| DEFINITION | nf36d10.s1 NCI_CGAP_Pri2 Homo sapiens cDNA clone IMAGE:915859, mRNA sequence. |

| | |
|-----------|------------|
| ACCESSION | AA579247 |
| NID | 92357431 |
| VERSION | AA579247.1 |
| | GI:2357431 |

| | |
|----------|--------------|
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |

| ORGANISM | REFERENCE |
|---|--|
| Homo sapiens | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| 1 (bases 1 to 387) | NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| NATIONAL CANCER INSTITUTE, CANCER GENOME ANATOMY PROJECT (CGAP), TUMOR GENE INDEX | Unpublished (1997) |

JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagut,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html

Insert Length: 681 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 351.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .387 |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915859"
/clone_lib="NCI_CGAP_pr2"
./sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/notes="vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

```

| | | | | |
|------------|-------|------|------|-------|
| BASE COUNT | 125 a | 54 c | 67 g | 141 t |
| ORIGIN | | | | |

| | | | | |
|-----------------------|--------|--------------------|--------|-------------|
| Query Match | 97.7%; | Score 374; | DB 35; | Length 387; |
| Best Local Similarity | 98.4%; | Pred. No. 1.3e-67; | | |

Matches 371; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATATCTTAGGACTGATAT 60
|||||
Db 11 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATATCTTAGGACTGATAT 70
61 TGGTATATATGTCATATTTAATWRTKRTKGGGCATTTCCCTTACATTCCTTGACAAGA 120
|||||
Db 71 TGGTATATATGTCATATTTAATATATTTTGGGGCATTTCCCTTACATTCCTTGACAAGA 130
QY 121 TTTAAATGCTGTGCCAAATTTTGTATTTATTTGGAGACTTCCTATTCACAAAGTAATGC 180
|||||
Db 131 TTTAAATGCTGTGCCAAATTTTGTATTTATTTGGAGACTTCCTATTCACAAAGTAATGC 190
QY 181 TGCCTAAGGAAGCTAAGGAATTAGTAGTGTCCCTCCTGCTGCTGCTATTC 240
|||||
Db 191 TGCCTAAGGAAGCTAAGGAATTAGTAGTGTCCCTCCTGCTGCTGCTATTC 250
QY 241 TAAAGATTTTGAATTCCTGGAATGACAAATATATTTTACTTTGGTGGGGAANAGTT 300
|||||
Db 251 TAAAGATTTTGAATTCCTGGAATGACAAATATATTTTACTTTGGTGGGGAANAGTT 310
QY 301 ATAGGACCACAGCTCTTCACTTCTGATACCTGTAAATTAATCTTTATTCGACTGTTTG 360
|||||
Db 311 ATAGGACCACAGCTCTTCACTTCTGATACCTGTAAATTAATCTTTATTCGACTGTTTG 370
QY 361 ACCATTAAGCTATATGT 377
|||||
Db 371 ACCATTAAGCTATATGT 387

RESULT 3
LOCUS W56075 522 bp mRNA EST 11-OCT-1996
DEFINITION zc56b05.r1 Soares_parathyroid_tumor_NbHRA Homo sapiens cdna clone
IMAGE:326289 5', mRNA sequence.
W56075
ACCESSION W56075
NID g1357965
VERSION W56075.1 GI:1357965
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Mairia, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785211.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 595 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 335.
Location/Qualifiers
1. 522

FEATURES
source
/organism="Homo sapiens"
/db_xref="GDB:1260491"
/db_xref="taxon:9606"
/map="643H05"
/clone="IMAGE:326289"
/clone_lib="Soares_parathyroid_tumor_NbHRA"

/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D
(pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cdna was primed with a Not I -
oligo(dt) primer
[5'-
TGTACCAATCTGAAGTGGGAGCGCGCCGACCAATTTTTTTTTTTTTTTTTT
T-3'], double-stranded cdna was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 171 a 71 c 94 g 181 t 5 others
ORIGIN
Query Match 96.6%; Score 369.8; DB 26; Length 522;
Best Local Similarity 97.9%; Pred. No. 8.6e-67;
Matches 368; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 7 AACAAACAAAAAACAATCTTCATTCAGAAAAATATATCTTAGGACTGATATGCTAA 66
|||||
Db 1 AACAAACAAAAAACAATCTTCATTCAGAAAAATATATCTTAGGACTGATATGCTAA 60
QY 67 TTATGTCATTTAATWRTKRTKGGGCATTTCCCTTACATTCCTTGACAAGATTAAGA 126
|||||
Db 61 TTATGTCATTTAATATATATTTTGGGCATTTCCCTTACATTCCTTGACAAGATTAAGA 120
QY 127 TGTCTGTGCCAAATTTGTATTTATTTGGAGACTTCTTATCAAAAGTATGCTGCCAA 186
|||||
Db 121 TGTCTGTGCCAAATTTGTATTTATTTGGAGACTTCTTATCAAAAGTATGCTGCCAA 180
QY 187 AGGAAGTCTAAGCAATTAGTAGTGTCCCTCCTGCTGCTGCTATTTAAAG 246
|||||
Db 181 AGGAAGTCTAAGCAATTAGTAGTGTCCCTCCTGCTGCTGCTATTTAAAG 240
QY 247 ATTTGATTTCTGGAATGACAAATTAATTTTAACTTTGGTGGGGAANAGTTATAGGA 306
|||||
Db 241 ATTTGATTTCTGGAATGACAAATTAATTTTAACTTTGGTGGGGAANAGTTATAGGA 300
QY 307 CCACAGTCTTCACTTCTGATACCTGTAAATTAATCTTTATTCGACTGTTTGACCAAT 366
|||||
Db 301 CCACAGTCTTCACTTCTGATACCTGTAAATTAATCTTTATTCGACTGTTTGACCAAT 360
QY 367 AAGCTATATGTTTAA 382
|||||
Db 361 AAGCTATATGTTTAA 376

RESULT 4
LOCUS A1188401/c 469 bp mRNA EST 28-OCT-1998
DEFINITION qd08e02.x1 Soares_placenta_8to9weeks_2NBHP8to9W Homo sapiens cdna
clone IMAGE:1723130 3', mRNA sequence.
A1188401
NID g3739610
VERSION A1188401.1 GI:3739610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 469)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1019 Std Error: 0.00

Seq primer: -40UP from Glibco

High quality sequence stop: 460.

FEATURES

source

1. .469
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1723130"

/clone_lib="Soares_placenta_8to9weeks_2NDHP8to9W"

/dev_stage="two placentae: one from 8 weeks and another

from 9 weeks post conception"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTACCAATCTGAAGTGGAGCGCGCGGATTTTCTTTTCTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaldo."

BASE COUNT 181 a 78 c 64 g 146 t

ORIGIN

Query Match 90.7%; Score 347.2; DB 43; Length 469;

Best Local Similarity 97.5%; Pred. No. 3.6e-62;

Matches 346; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 26 TCCTCATTCAGAAAAATATCTTAGGGAGCTGATTTGGTAATTAGGTCATTTAATWRT 85

Db 469 TCCTCATTCAGAAAAATATCTTAGGGAGCTGATTTGGTAATTAGGTCATTTAAT 410

QY 86 RTTGTGGGCGATTTCCCTTACATTTGCTTGACAAGATTAAATGCTGTGCCAAATTTTG 145

Db 409 ATTTGGGGCATTTCCCTTACATTTGCTTGACAAGATTAAATGCTGTGCCAAATTTTG 350

QY 146 TATTTTATTTGAGACCTTCTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAG 205

Db 349 TATTTTATTTGAGACCTTCTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAG 290

QY 206 TAGTGTCCCMTCACCTGTGTTGGAGTGTGCTATTTCTAAAGATTTTGATTTCCCTGAATG 265

Db 289 TAGTGTCCCMTCACCTGTGTTGGAGTGTGCTATTTCTAAAGATTTTGATTTCCCTGAATG 230

QY 266 ACAATTATATTTTAACTTTGGTGGGGAANAGTTATAGGACCACAGCTTCTCACTCTGA 325

Db 229 ACAATTATATTTTAACTTTGGTGGGGAANAGTTATAGGACCACAGCTTCTCACTCTGA 170

QY 326 TACTGTAAATTAATCTTTTATTTGACCTGTTTGAACCATTTAGCTATATGTTTA 380

Db 169 TACTGTAAATTAATCTTTTATTTGACCTGTTTGAACCATTTAGCTATATGTTTA 115

RESULT 5

LOCUS

C75566

DEFINITION

C75566

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 432)
AUTHORS Takeda, J.
TITLE Large scale collection of expressed sequence tags (ESTs) from human
pancreatic islet cDNA library
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:804206.

Contact: Jun Takeda
Institute for Molecular and Cellular Regulation, Gunma University
3-39-15 Showa-machi, Maebashi Gunma 371, Japan
Tel: 272-20-8856
Fax: 272-20-8896
Email: jtakeda@eb.gunma-u.ac.jp.

FEATURES

source

1. .432
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="hbc7876"

/clone_lib="Human pancreatic islet"

/note="Vector: Lambda ZAP11; Site_1: Eco RI; Site_2: Xho

I; mRNA was prepared from normal adult human islets. cDNA

was directionally synthesized from the Xho I in the vector

to the EcoRI site. cDNA was size fractionated to remove

sequences <1000 bp in size."

BASE COUNT 163 a 79 c 52 g 137 t 1 others

ORIGIN

Query Match 88.6%; Score 339.4; DB 35; Length 432;

Best Local Similarity 97.7%; Pred. No. 1.5e-60;

Matches 337; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 39 AAATATCTTAGGGAGCTGATATTGGTAATATGCTCAATTTAAATWRTTKTGGGGCAT 98

Db 432 AAATATCTTAGGGAGCTGATATTGGTAATATGCTCAATTTAAATATATTTGGGGCAT 373

QY 99 TCCTTACATTTGCTTGACAAGATTAAATGCTGTGCCAAATTTGTAATTTATTTGGA 158

Db 372 TCCTTACATTTGCTTGACAAGATTAAATGCTGTGCCAAATTTGTAATTTATTTGGA 313

QY 159 GACTTCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAGTGTCCCMTC 218

Db 312 GACTTCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAGTGTCCCMTC 253

QY 219 ACTTGTGAGTGTGCTATTTCTAAAGATTTTGATTTCCCTGAATGACATATATTTT 278

Db 252 ACTTGTGAGTGTGCTATTTCTAAAGATTTTGATTTCCCTGAATGACATATATTTT 193

QY 279 AACTTGTGAGGGAANAGTTATAGGACCACAGCTTCTCACTTCTGATCTGTAATTA 338

Db 192 AACTTGTGAGGGAANAGTTATAGGACCACAGCTTCTCACTTCTGATCTGTAATTA 133

QY 339 ATCTTTATTTGACCTGTTTGTGACCATTTAGCTATATGTTTAAA 383

Db 132 ATCTTTATTTGACCTGTTTGTGACCATTTAGCTATATGTTTAAA 88

RESULT 6

LOCUS

AI027536

DEFINITION

AI027536

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)

NID

SOURCE

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1899479.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert length: 564 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 444.

FEATURES

source

1.472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1650448"
/clone_lib="Soares_parathyroid_tumor_NBHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
15'-

TGTTACCAATCTGAAGTGGAGCGCGCACCAATTTTCTTTTCTTTTCTTTT
T-3'}, double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 175 a 80 c 55 g 162 t
ORIGIN

Query Match 87.3%; Score 334.4; DB 41; Length 472;
Best Local Similarity 97.9%; Pred. No. 1.5e-59;
Matches 332; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 45 TCTTAGGAGCTGATATGTGAATTAATGTCATTTAAATWRTTKTGGGCAATTCCTTA 104
|||
DB 472 TCTTAGGAGCTGATATGTGAATTAATGTCATTTAAATATTTTGGGCAATTCCTTA 413
OY 105 CATGCTTGACAGATTAATAATGCTGTGCCAAAATTTGTATTTATTTGGAGACTTC 164
|||
DB 412 CATGCTTGACAGATTAATAATGCTGTGCCAAAATTTGTATTTATTTGGAGACTTC 353
OY 165 TTATCAAAAGTATGCTGCCAAGAGAGTCTAAGGAATTAAGTAGTGTCCCMTCACCTGT 224
|||
DB 352 TTATCAAAAGTATGCTGCCAAGAGAGTCTAAGGAATTAAGTAGTGTCCCMTCACCTGT 293
OY 225 TTGGAGTGTCTATCTTAAGATTTTGTATTCCTGGAATGACAATTAATTTAACTTT 284
|||
DB 292 TTGGAGTGTCTATCTTAAGATTTTGTATTCCTGGAATGACAATTAATTTAACTTT 233
OY 285 GGTGGGGAANAGTTATAGACACAGCTCTTCACTTCGTATCTTGAATTAATCTTT 344
|||
DB 232 GGTGGGGAANAGTTATAGACACAGCTCTTCACTTCGTATCTTGAATTAATCTTT 173
OY 345 TATTGCACTTGTGACCATTAAGCTATATGTTAA 383
|||

DB 172 TATTGCACTTGTGACCATTAAGCTATATGTTAGAA 134

RESULT 7
AA879150 389 bp mRNA EST 25-MAR-1998
LOCUS
DEFINITION nw84e05.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1253312,
mRNA sequence.
ACCESSION AA879150
NID 92988115
VERSION AA879150.1 GI:2988115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 389)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153434.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Douglas Fligg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 381.

FEATURES

source

1.389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone="IMAGE:1253312"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

BASE COUNT 121 a 51 c 75 g 142 t
ORIGIN

Query Match 85.5%; Score 327.4; DB 39; Length 389;
Best Local Similarity 95.5%; Pred. No. 4.2e-58;
Matches 342; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

OY 26 TCTTCATTCAGAAAAATATCTTAGGAGCTGATTTGTAATTAATGTCATTAATWRT 85
|||
DB 3 TCGACATTCAGAAAAATATCTTAGGAGCTGATTTGTAATTAATGTCATTAATTAAT 62
OY 86 RTTKTGGGCAATTCCTTACATTTGCTTGACAGATTAATGTCGTGCCAAATTTTG 145
|||
DB 63 ATT-TGGGGCAATTCCTTACATTTGCTTGACAGATTAATGTCGTGCCAAATTTTG 121
OY 146 TATTTTATTTGGAGACTTCTATCAAAAGTAATGCTGCCAAAGAGAGCTTAAGGAATTAG 205
|||
DB 122 TATTTTATTTGGAGACTTCTATCAAAAGTAATGCTGCCAAAGAGAGCTTAAGGAATTAG 181
OY 206 TAGTGTCCCMTCACCTGTTTGAGAGTGTCTAATCTAAAGATTTGATTTCTCTGAATG 265
|||

|||||:|||||
Db 182 TAGGTCCACACACTGTTGGAGTGTGCAATCTAAAGATTGATTTCTCGAATG 241
QY 266 ACAATTATATTTAACTTTGGTGGGGAANAAGTTATAGACACAGCTTCACTTCTGA 325
|||||
Db 242 ACAATTATATTTAACTTTGGTGGGGAAGAGTTATAGACACAGCTTCACTTCTGA 301
QY 326 TACTTGTAAATTAATCTTTTATGTCACCTGTTTGACCATTAAGCTATATGTTAAAA 383
|||||
Db 302 TACTTGTAAATTAATCTTTTATGTCACCTGTTTGACCATTAAGCTATATGTTAGAA 359
RESULT 8
AI160743/c 442 bp mRNA EST 26-OCT-1998
LOCUS gb49d03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1703429 3',
DEFINITION mRNA sequence.
ACCESSION AI160743
NID 93694123
VERSION AI160743.1 GI:3694123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 442)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute / National Institute of Neurological
TITLE Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286564.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2616 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 433.
Location/Qualifiers
1. .442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1703429"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; site_1: Not I; site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 172 a 76 c 53 g 141 t
ORIGIN

Query Match 83.7%; Score 320.6; DB 43; Length 442;
Best Local Similarity 97.0%; Pred. No. 1e-56;
Matches 320; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 54 CTGATATTGGTAATTATGTCAAATTAAWTRTRTKTGGGCATTTCCCTACATGTCTT 113
|||||
Db 442 CTGATATTGGTAATTATGTCAAATTAAATATATTTGAGGCATTTCCCTACATGTCTT 383
QY 114 GACAAGATTAATAATGCTGTGCCAANAATTTGTATTTTATTTGGAGACTTCTATCAAAA 173
|||||
Db 382 GACAAGATTAATAATGCTGTGCCAACAATTTGTATTTTATTTGGAGACTTCTATCAAAA 323
QY 174 GTATGCTGCCAAGGAAGTCTAAGCAATTAGTAGTGTTCCTCCMTCACTGTGTTGGAGTGT 233
|||||
Db 322 GTATGCTGCCAAGGAAGTCTAAGCAATTAGTAGTGTTCCTCCATCACTGTGTTGGAGTGT 263
QY 234 GCTATCTAAAGATTTTGATTTCCCTGGAAATGACAATTATATTTTAACCTTTGGTGGGGA 293
|||||
Db 262 GCTATCTAAAGATTTTGATTTCCCTGGAAATGACAATTATATTTTAACCTTTGGTGGGGA 203
QY 294 AANAGTTATAGCACACAGCTCTTCACTCTGATACCTGTGAATTAATCTTTATGCACT 353
|||||
Db 202 AAGAGTTATAGCACACAGCTCTTCACTCTGATACCTGTGAATTAATCTTTATGCACT 143
QY 354 TGTTTGACCATTAAGCTATATGTTAAAA 383
|||||
Db 142 TGTTTGGCCATTAAAGCTATATGTTAGAA 113
RESULT 9
AI522136/c 444 bp mRNA EST 13-APR-1999
LOCUS t178f03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2138141 3',
DEFINITION mRNA sequence.
ACCESSION AI522136
NID 94436271
VERSION AI522136.1 GI:4436271
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948750.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 539 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2138141"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; site_1: Not I; site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs

FEATURES
source

from a pool of 5,000 clones made from the same library
(clonids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo.

BASE COUNT 170 a 75 c 51 g 148 t
ORIGIN

Query Match 82.8%; Score 317.4; DB 47; Length 444;
Best Local Similarity 97.8%; Pred. No. 4.5e-56;
Matches 315; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 62 GGTATTATGTCATTTATATWRTTRTKGGGCATTTCCCTACATTTGCTTGACAAGAT 121
|||||
Db 444 GGTATTATGTCATTTATATATATTTGGGCATTTCCCTACATTTGCTTGACAAGAT 385
QY 122 TAAATGTCGTGCCAAAATTTGTATTTTGGAGACTTCTTATCAAAAGTAATGCT 181
|||||
Db 384 TAAATGTCGTGCCAAAATTTGTATTTTGGAGACTTCTTATCAAAAGTAATGCT 325
QY 182 GCCAAGAGACTTAAGGAATTAGTAGTCCCTCCTGAGAGTGTGCTATCT 241
|||||
Db 324 GCCAAGAGAGCTAAGGAATTAGTAGTCCCTCCTGAGAGTGTGCTATCT 265
QY 242 AAAAGATTTGATTTCTCGAATGACAAATATTTTAACTTTGGTGGGGAANAGTTA 301
|||||
Db 264 AAAAGATTTGATTTCTCGAATGACAAATATTTTAACTTTGGTGGGGAANAGTTA 205
QY 302 TAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATTTGACACTTTTGA 361
|||||
Db 204 TAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATTTGACACTTTTGA 145
QY 362 CCATTAGCTATATGTTTAAA 383
|||||
Db 144 CCATTAGCTATATGTTTAAA 123

RESULT 10
LOCUS C75584/c 419 bp mRNA EST 09-SEP-1997
DEFINITION C75584 Human pancreatic islet Homo sapiens cDNA clone hbc7897, mRNA
sequence.
ACCESSION C75584
NID 92366646
VERSION C75584.1 GI:2366646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 419)
Takeda, J.
TITLE Large scale collection of expressed sequence tags (ESTs) from human
pancreatic islet cDNA library
JOURNAL Unpublished (1997)
COMMENT On May 5, 1995 this sequence version replaced gi:797622.

CONTACT: Jun Takeda
Institute for Molecular and Cellular Regulation, Gunma University
3-39-15 Showa-machi, Maebashi Gunma 371, Japan
Tel: 272-20-8856
Fax: 272-20-8896
Email: jtakeda@sb.gunma-u.ac.jp.
LOCATION/Qualifiers
1. 419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hbc7897"
/clone_lib="Human pancreatic islet"
/note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
I; mRNA was prepared from normal adult human islets. cDNA
was directionally synthesized from the Xho I in the vector
to the EcoRI site. cDNA was size fractionated to remove

sequences <1000 bp in size."
BASE COUNT 159 a 76 c 50 g 132 t 2 others
ORIGIN

Query Match 81.8%; Score 313.4; DB 35; Length 419;
Best Local Similarity 97.0%; Pred. No. 3e-55;
Matches 322; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 53 ACTGATATGCTAATATATGTCATTTATWRTTRTKGGGCATTT-CCCTACATGTC 111
|||||
Db 419 ACTGATATGCTAATATATGTCATTTATATATATTTGGGCATTTCCCTACATGTC 360
QY 112 TTGACAAGATTAAATGCTCTGCCAAAATTTGTATTTTATTTGGAGACTTATCAA 171
|||||
Db 359 TTGACAAGATTAAATGCTCTGCCAAAATTTGTATTTTATTTGGAGACTTATCAA 300
QY 172 AAGTAATGCTGCCAAGAGCTAAGGAATTAGTAGTCCCTCCTGAGAGT 231
|||||
Db 299 AAGTAATGCTGCCAAGAGCTAAGGAATTAGTAGTCCCTCCTGAGAGT 240
QY 232 GTGCTATTTAAAGATTTTGAATTTCTCGAATGACAAATATTTTAACTTTGGTGGG 291
|||||
Db 239 GTGCTATTTAAAGATTTTGAATTTCTCGAATGACAAATATTTTAACTTTGGTGGG 180
QY 292 GAANAGTTATAGGACACAGCTTCTCACTTCTGATACCTGTAATTAATCTTTATGCA 351
|||||
Db 179 GAAGAGCTTATAGGACACAGCTTCTCACTTCTGATACCTGTAATTAATCTTTATGCA 120
QY 352 CTGTTTGGACCATTTAAGCTATATGTTTAAA 383
|||||
Db 119 CTGTTTGGACCATTTAAGCTATATGTTTAAA 88

RESULT 11
LOCUS AI005274/c 496 bp mRNA EST 27-AUG-1998
DEFINITION ou07f12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1625615 3', mRNA sequence.
ACCESSION AI005274
NID 93214784
VERSION AI005274.1 GI:3214784
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1636 Std Error: 0.00
Seq primer: primer name ambiguous
High quality sequence stop: 454.
LOCATION/Qualifiers
1. 496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1625615"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbH19w, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 193 a 79 c 57 g 167 t
ORIGIN

Query Match 79.2%; Score 303.4; DB 41; Length 496;
Best Local Similarity 97.5%; Pred. No. 3.2e-53;
Matches 312; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

QY 64 TAATATGTCATTTAAWTRTKTGGGCATTCTTACATGTCGTGACAAATTA 123
|||||
Db 496 TAATATGTCATTTAAATATATTTGGGGCATTCTTACATGTCGTGACAA-ATTA 438
QY 124 AAATGCTCTGCCAAAATTTGTATTATTGGAGACTTCTATCAAAAGTAATGCTGC 183
|||||
Db 437 AAATGCTCTGCCAAAATTTGTATTATTGGAGACTTCTATCAAAAGTAATGCTGC 378
QY 184 CAAGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTTGGAGTGCCTATTCTAA 243
|||||
Db 377 CAAGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTTGGAGTGCCTATTCTAA 318
QY 244 AAGATTTTGATTTCCGTGAATGACAAATATATTAACTTTGTTGGGGAAANAGTTATA 303
|||||
Db 317 AAGATTTTGATTTCCGTGAATGACAAATATATTAACTTTGTTGGGGAAAGATTATA 258
QY 304 GGACCAACAGTCTTCACCTTCGATACTGTAAATTAATCTTTATTCGACCTGTTTGACC 363
|||||
Db 257 GGACCAACAGTCTTCACCTTCGATACTGTAAATTAATCTTTATTCGACCTGTTTGACC 198
QY 364 ATTAAGCTATATGTTTAAA 383
|||||
Db 197 ATTAAGCTATATGTTTAAA 178

RESULT 12
AI051146 473 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION OY49604.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1669159 3',
mRNA sequence.

ACCESSION AI051146
NID 93306680
VERSION AI051146.1 GI:3306680
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
1 (bases 1 to 473)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGA), Tumor Gene Index
Unpublished (1998)

JOURNAL
COMMENT
On Jan 17, 1998 this sequence version replaced gi:2044831.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/bdrip/image/image.html

FEATURES
source
Insert Length: 2319 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 438.
Location/Qualifiers
1. .473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1669159"
/clone_lib="NCI_CGAP_Brn23"
/issue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCCGCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 181 a 77 c 56 g 159 t
ORIGIN

Query Match 76.1%; Score 291.6; DB 41; Length 473;
Best Local Similarity 98.6%; Pred. No. 8.4e-51;
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 89 KTGCGGCATTTCCTACATGTCTTGACACAGATTAAATGTCGTGCCAAAATTTGTAT 148
:|||||
Db 473 TTGGGGCATTTCCCTACATGTCTTGACACAGATTAAATGTCGTGCCAAAATTTGTAT 414
QY 149 TTATTTTGAGACTCTTATCAAAAGTAATGCTGCCAAAGAGTCTAAGGAATTAGTAG 208
|||||
Db 413 TTATTTTGAGACTCTTATCAAAAGTAATGCTGCCAAAGAGTCTAAGGAATTAGTAG 354
QY 209 TGTCCCMTCACCTGTTGGAGTGTGCTAATCTTAAAGATTTGATTTCCCTGGAATGACA 268
|||||
Db 353 TGTCCCMTCACCTGTTGGAGTGTGCTAATCTTAAAGATTTGATTTCCCTGGAATGACA 294
QY 269 ATTATATTTTAACCTTGGTGGGGAANAAGTTATAGACACACAGTCTTCACTTCTGATAC 328
|||||
Db 293 ATTATATTTTAACCTTGGTGGGGAANAAGTTATAGACACACAGTCTTCACTTCTGATAC 234
QY 329 TTGTAATTAATCTTTATTTGCACTGTTTGGACCATTAAGCTATATGTTTAAA 383
|||||
Db 233 TTGTAATTAATCTTTATTTGCACTGTTTGGACCATTAAGCTATATGTTTAAA 179

RESULT 13
AA622892 407 bp mRNA EST 21-OCT-1997
LOCUS
DEFINITION np58c12.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1130518 3',
similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA622892
NID 92526768
VERSION AA622892.1 GI:2526768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
1 (bases 1 to 407)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT
On May 18, 1995 this sequence version replaced gi:810969.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/abbrp/image/image.html

Insert Length: 948 Std Error: 0.00
Seq. primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 385.

FEATURES

source

1. 407

location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1130518"

/clone_lib="NCI-CGAP_Br2"

/sex="female, pooled"

/tissue_type="breast"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. This library is the normalized version of NCI-CGAP_Br1.1. Library was constructed by Bento Soares and M. Fatima Ronaldo."

BASE COUNT 157 a 72 c 50 g 128 t
ORIGIN

Query Match

Best Local Similarity 75.2%; Score 288; DB 36; Length 407;

Matches 288; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 93 GGCATTTCTTACATTTCTTGACAGATTAAATGTCGTGCCAAATTTGTATTTTA 152
DB 407 GGCATTTCTTACATTTCTTGACAGATTAAATGTCGTGCCAAATTTGTATTTTA 348
OY 153 TTGGAGACTTCTATCAAAAGTATGCTGCCAAGAGTCTAAGGAATTAGTAGTGT 212
DB 347 TTGGAGACTTCTATCAAAAGTATGCTGCCAAGAGTCTAAGGAATTAGTAGTGT 288
OY 213 CCCATCAGTTGTTGAGTGTGCTATTTCTAAAGATTGATTTCTCTGGAATGACATTA 272
DB 287 CCCATCAGTTGTTGAGTGTGCTATTTCTAAAGATTGATTTCTCTGGAATGACATTA 228
OY 273 TATTTTAAGTTGTTGGGGAANAAGTTATAGACACAGTCTTCACTTCTGATACTTGT 332
DB 227 TATTTTAAGTTGTTGGGGAAGAGTTATAGACACAGTCTTCACTTCTGATACTTGT 168
OY 333 AAATTAATCTTTTATGACACTTGTGACCATTAAGCTATATGTTTAAA 383
DB 167 AAATTAATCTTTTATGACACTTGTGACCATTAAGCTATATGTTTAAA 117

RESULT 14

C14656

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

344 bp mRNA EST 30-SEP-1996
C14656 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens CDNA
clone GEN-078E02 5', mRNA sequence.
C14656
91569363
C14656.1 GI:1569363
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 344)

AUTHORS

Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinozumiya,H.,
Takahachi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
Maekawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
Unpublished (1995)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407498.

CONTACT: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
463-10 kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES

source

location/Qualifiers

1. 344

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GEN-078E02"

/clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT

138 a 44 c 50 g 108 t 4 others

ORIGIN

Query Match

Best Local Similarity 73.2%; Score 280.4; DB 28; Length 344;

Matches 276; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAAAAACAATTCCTCATTCAGAAAAATTATCTTAGGAGCTGATAT 60
DB 59 AAACAACAACAACAAAAACAATTCCTCATTCAGAAAAATTATCTTAGGAGCTGATAT 118
OY 61 TGGTATATGTCATTAATWRTTGTGGGCAATTCCTTACATGCTTGACAAGA 120
DB 119 TGGTATATGTCATTAATWRTTGTGGGCAATTCCTTACATGCTTGACAAGA 178
OY 121 TTAATATGCTGTGCCAATAATTTGTATTTATTTGGAGACTTCTTAACAAGTATGC 180
DB 179 TTAATATGCTGTGCCAATAATTTGTATTTATTTGGAGACTTCTTAACAAGTATGC 238
OY 181 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC 240
DB 239 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC 298
OY 241 TAAAGATTTGATTTCTCGAATGACAAATTAATTTAATTTTAACTTTGG 286
DB 299 TAAAGATTTGATTTCTCGAATGACAAATTAATTTAATTTTCAACTTTGG 344

RESULT 15

A1420653/c

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

463 bp mRNA EST 28-MAR-1999
A1420653 tef13b04.x1 NCI-CGAP_Brn23 Homo sapiens CDNA clone IMAGE:2096047 3',
mRNA sequence.
A1420653
94266584
A1420653.1 GI:4266584
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 463)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGAP), Tumor Gene Index
Unpublished (1998)
On Mar 20, 1998 this sequence version replaced gi:2980399.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1420 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 449.
Location/Qualifiers

FEATURES

source

1. .463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2096047"
/clone_lib="NCI-CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGACATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 179 a 76 c 55 g 153 t
ORIGIN

Query Match 72.38; Score 277; DB 46; Length 463;
Best Local Similarity 98.68; Pred. No. 8.1e-48;
Matches 288; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

OY 92 GGGCATTTCCTTACATGCTTGACAAGATTAATGCTGTGCCAAATTTGTATTTT 151
|||||
Db 463 GGGCATTTCCTTACATGCTTGACAAGATTAATGCTGTGCCAAATTTGTATTTT 404
OY 152 ATTTGAGACTTCTATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAGTAGT 211
|||||
Db 403 ATTTGAGACTT-TTATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAGTAGT 345
OY 212 TCCCMTCACCTGTTGGAGTGTGCTATCTTAAGAATTTGATTTCCCTGGAATGACAATT 271
|||||
Db 344 TCCCATCAGCTGTTGGAGTGTGCTATCTTAAGAATTTGATTTCCCTGGAATGACAATT 285
OY 272 ATATTTTAACTTTGGTGGGGGAAANAGTTATAGACCAAGCTCTTCACTTCTGATACCTTG 331
|||||
Db 284 ATATTTTAACTTTGGTGGGGGAAAGAGTTATAGACCACAGTCTTCACTTCTGATACCTTG 225
OY 332 TAAATTAATCTTTTATGCACTGTTTGGACCAATTAAGCTATATGTTTAAA 383
|||||
Db 224 TAAATTAATCTTTTATGCACTGTTTGGACCAATTAAGCTATATGTTTAAA 173

Search completed: September 28, 1999, 11:29:54
Job time: 1875 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:40:51 ; Search time 2910.9 Seconds
(without alignments)
418.447 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383

Sequence: 1 AAAACAACAACAACAAAAA.....ATTACCTATATGTTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: *
1: gb_bal: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vl: *
18: em_fun: *
19: em_htg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vl: *
34: gb_htg1: *
35: gb_htg2: *
36: gb_in1: *
37: gb_in2: *
38: em_bal: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pr4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result 8
No. Score Query Match Length DB ID Description

C 1 378.4 98.8 197441 34 CNS00005 AL049779 Homo sapi

| | | | | | | | |
|---|----|------|------|--------|----|------------|--------------------|
| C | 2 | 45.6 | 11.9 | 149764 | 10 | HSAC002076 | AC002076 Human BAC |
| | 3 | 45.2 | 11.8 | 207139 | 34 | CEY17G7 | Z96049 Caenorhabd1 |
| | 4 | 45.2 | 11.8 | 143092 | 36 | CEY17G7B | AL023828 Caenorhab |
| C | 5 | 45 | 11.7 | 104308 | 8 | AC006228 | AC006228 Genomic s |
| C | 6 | 44.6 | 11.6 | 82044 | 34 | PFMAL1P2 | AL031745 Plasmodi |
| | 7 | 44.6 | 11.6 | 441 | 36 | IPU06960 | U06960 Ichneumon p |
| | 8 | 44.2 | 11.5 | 151798 | 9 | HS435C23 | Z92844 Human DNA s |
| C | 9 | 44 | 11.5 | 172007 | 42 | AC006376 | AC006376 Homo sapi |
| | 10 | 43.8 | 11.4 | 110680 | 8 | AC006259 | AC006259 Arabidops |
| | 11 | 43.8 | 11.4 | 36823 | 10 | HSU209G1 | Z68873 Human DNA s |
| | 12 | 43.8 | 11.4 | 216991 | 34 | HS164F3 | AL035422 Homo sapi |
| | 13 | 43.2 | 11.3 | 4284 | 36 | DDU20661 | U20661 Dictyoste11 |
| C | 14 | 42.8 | 11.2 | 96642 | 8 | AC002291 | AC002291 Arabidops |
| C | 15 | 42.8 | 11.2 | 92612 | 8 | ATAC003974 | AC003974 Arabidops |
| C | 16 | 42.8 | 11.2 | 274690 | 34 | CEY60A9 | AL022281 Caenorhab |
| C | 17 | 42.8 | 11.2 | 167846 | 34 | CEY66C5 | Z98874 Caenorhabd1 |
| C | 18 | 42.8 | 11.2 | 24598 | 36 | CET24C2 | Z68120 Caenorhabd1 |
| C | 19 | 42.8 | 11.2 | 12029 | 37 | AE001400 | AE001400 Plasmodi |
| | 20 | 42.6 | 11.1 | 300172 | 35 | AC005308 | AC005308 Plasmodi |
| C | 21 | 42.4 | 11.1 | 143751 | 11 | AC004160 | AC004160 Homo sapi |
| C | 22 | 42.4 | 11.1 | 142573 | 34 | HSUJ753D5 | AL049693 Homo sapi |
| C | 23 | 42.4 | 11.1 | 929 | 36 | DDIDDCOF1 | D37980 Dictyoste11 |
| C | 24 | 42.4 | 11.1 | 3946 | 36 | DDU25143 | U25143 Dictyoste11 |
| C | 25 | 42 | 11.0 | 143299 | 10 | HS81H13 | AL023805 Human DNA |
| C | 26 | 42 | 11.0 | 199606 | 35 | AC004688 | AC004688 Plasmodi |
| C | 27 | 42 | 11.0 | 194410 | 35 | AC005140 | AC005140 Plasmodi |
| C | 28 | 41.8 | 10.9 | 89281 | 8 | ATAC006304 | AC006304 Arabidops |
| | 29 | 41.8 | 10.9 | 97494 | 8 | T15F16 | AF076275 Arabidops |
| | 30 | 41.8 | 10.9 | 99902 | 10 | HS59B16 | AL032822 Human DNA |
| C | 31 | 41.8 | 10.9 | 86829 | 36 | PFMAL3P5 | AL034556 Plasmodi |
| C | 32 | 41.6 | 10.9 | 74342 | 7 | AB020742 | AB020742 Arabidops |
| C | 33 | 41.4 | 10.8 | 107331 | 34 | CEY49A10 | Z93240 Caenorhabd1 |
| C | 34 | 41.4 | 10.8 | 312766 | 34 | PFMAL4P3 | AL035476 Plasmodi |
| C | 35 | 41.4 | 10.8 | 37750 | 36 | CEZC504 | Z50029 Caenorhabd1 |
| | 36 | 41.2 | 10.8 | 12039 | 3 | BTU25810 | U25810 Bos taurus |
| | 37 | 41.2 | 10.8 | 79574 | 8 | ATAC002340 | AC002340 Arabidops |
| C | 38 | 41.2 | 10.8 | 156909 | 9 | AB020863 | AB020863 Homo sapi |
| C | 39 | 41.2 | 10.8 | 146813 | 11 | AF121898 | AF121898 Homo sapi |
| C | 40 | 41 | 10.7 | 299081 | 34 | AC006892 | AC006892 Caenorhab |
| C | 41 | 41 | 10.7 | 40168 | 36 | CELR06B10 | AF040654 Caenorhab |
| C | 42 | 40.8 | 10.7 | 83450 | 7 | AB009054 | AB009054 Arabidops |
| C | 43 | 40.8 | 10.7 | 82594 | 8 | ATAC005311 | AC005311 Arabidops |
| C | 44 | 40.8 | 10.7 | 154848 | 11 | AC002992 | AC002992 Homo sapi |
| C | 45 | 40.8 | 10.7 | 160262 | 11 | AC005242 | AC005242 Homo sapi |

ALIGNMENTS

RESULT 1
CNS00005/c
LOCUS
DEFINITION
Homo sapiens chromosome 14 clone bac R-1012A1 from RPCI-11 library,
WORKING DRAFT SEQUENCE, in ordered pieces.

ACCESSION
AL049779
NID
94760257
VERSION
AL049779.1 GI:4760257
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 197441)
AUTHORS
Direct Submission
TITLE
Submitted (05-MAY-1999) Genoscope - Centre national de sequence
JOURNAL
2, rue Gaston Cremieux - BP 191 91006 EVRY cedex - FRANCE E-mail :

COMMENT

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

Location/Qualifiers

source 1..197441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="bac R-1012A1 from RPCI-11 library"
BASE COUNT 57409 a 42215 c 42060 g 55757 t
ORIGIN

Query Match 98.8%; Score 378.4; DB 34; Length 197441;
Best Local Similarity 98.2%; Pred. No. 5.3e-61;
Matches 376; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 60
Db 50969 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATATCTTAGGACTGATAT 50910
QY 61 TGGTAATTATGGTCAATTAAATWRTTKTGGGGCAATTCCTTACATTTGCTTGACAAGA 120
Db 50909 TGGTAATTATGGTCAATTAAATATATTTTGGGGCAATTCCTTACATTTGCTTGACAAGA 50850
QY 121 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGC 180
Db 50849 TTAATAATGCTGTGCCAAAAATTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGC 50790
QY 181 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTTGTGGAGTGTCTATTTC 240
Db 50789 TGCCAAGGAAGTCTAAGGAATTAGTAGTGTCCCMTCACCTTGTGGAGTGTCTATTTC 50730
QY 241 TAAAGATTTTGTATTCCTGGAATGACATATATATTTTAACCTTGTGGGGGAANAGTT 300
Db 50729 TAAAGATTTTGTATTCCTGGAATGACATATATATTTTAACCTTGTGGGGGAANAGTT 50670
QY 301 ATAGACCACAGTCTTCACCTCTGATACCTGTGAATTAATCTTTATTCGACTGTTTG 360
Db 50669 ATAGACCACAGTCTTCACCTCTGATACCTGTGAATTAATCTTTATTCGACTGTTTG 50610
QY 361 ACCATTAAAGCTATATGTTTAAA 383
Db 50609 ACCATTAAAGCTATATGTTTAAA 50587

RESULT 2
HSAC002076/c HSAC002076 149764 bp DNA PRI 12-MAY-1997
LOCUS
DEFINITION Human BAC clone GS345D13 from 7q31-q32, complete sequence.
ACCESSION AC002076
NID g2078461
VERSION AC002076.1 GI:2078461
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149764)
AUTHORS Maggi, L.
TITLE The sequence of H. sapiens BAC clone GS345D13
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 149764)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1997)
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/MTB/CHR7> or send an E-mail to egreen@nhgri.nih.gov
Mapping information for this clone was also provided by Dr. John D. McPherson, Department of Genetics/Genome Sequencing Center, Washington University School of Medicine.

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelOBAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is H_GS274A07, 200 bp overlap. Actual start of this clone is at base position 1 of H_GS345D13; actual end is at 19827 of H_GS274A07. This cosmid lies in an unanchored cluster of unknown orientation.

This clone contains STS SWSS4055 (NID:g1916534), SWSS2840 (NID:g1113614), and SWSS1647 (NID:g1113169).

FEATURES
source 1..149764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="GS345D13"
/map="7q31-q32"
/map="7q31-q32"
3138..4124
/rpt_family="L1"
4162..4237
/rpt_family="L1"
complement(4345..4638)
/rpt_family="ALU"
4769..5162
/note="match to human EST AA039321 (NID:g1515599)
2k39604.s1"
complement(5119..5551)
/note="match to human EST AA039422 (NID:g1515699)
2k39604.r1"
6029..6148
/rpt_family="L1"
6787..7179
/note="match to human EST AA150894 (NID:g1722424)
z147c03.s1"
6788..7175
/note="match to human EST AA135883 (NID:g1696857)
z122b10.s1"
6788..7201
/note="match to human EST N33317 (NID:g1153716)
yy08a02.s1"
6802..7102
/note="match to human EST H01232 (NID:g864165) yj31f03.s1"
6909..7179
/note="match to human EST H42319 (NID:g918371) yc009g04.s1"
complement(7074..7211)


```

misc_feature      /note="match to human EST W34968 (NID:g1316887)
                  mc33e06.r1"
                  complement(7074. .7211)
                  /note="match to human EST AA030363 (NID:g1497501)
                  m11c04.r1"
                  complement(7074. .7211)
                  /note="match to human EST W45750 (NID:g1330484)
                  mc79g10.r1"
                  complement(7084. .11162)
                  /gene="WUGSC:GS345D13.3"
                  complement(join(7084. .7209,11067. .11162))
                  /gene="WUGSC:GS345D13.3"
                  /note="GS345D13.3; match to protein P50152 (PID:g1730223)"
                  /codon_start=1
                  /product="G-protein gamma-11 subunit"
                  /protein_id="AAB54051.1"
                  /db_xref="PID:g2078464"
                  /db_xref="GI:2078464"
                  /translation="MPALHIEDLPKEKLMKEVEQLRKEVKLQRQVSKSSEIKNYI
                  EERSGEDPVPKGIPEDKNPFKEKGCVIS"
                  EERSGEDPVPKGIPEDKNPFKEKGCVIS"
                  7993. .8035
repeat_region     /rpt_family="L1"
                  /complement(8127. .8417)
repeat_region     /rpt_family="ALU"
                  /complement(11064. .11181)
                  /note="match to human EST AA030363 (NID:g1497501)
                  m11c04.r1"
                  /complement(11064. .11181)
                  /note="match to human EST W45750 (NID:g1330484)
                  mc79g10.r1"
                  /complement(11064. .11181)
                  /note="match to human EST W34968 (NID:g1316887)
                  mc33e06.r1"
                  /complement(11065. .11236)
                  /note="match to human EST H00850 (NID:g863783) YJ31f03.r1"
                  /complement(11065. .11383)
                  /note="match to human EST AA151020 (NID:g1722531)
                  z147c03.r1"
                  /complement(11122. .11519)
                  /note="match to human EST N42976 (NID:g1166720)
                  YY08a02.r1"
                  /complement(11137. .11476)
                  /note="match to human EST AA135882 (NID:g1696856)
                  z122b10.r1"
                  /complement(13202. .13382)
                  /rpt_family="ALU"
                  /complement(13681. .13971)
                  /rpt_family="ALU"
                  /complement(14025. .14051)
                  /rpt_family="L1"
                  14675. .14792
repeat_region     /rpt_family="ALU"
                  /complement(15009. .15303)
                  /rpt_family="ALU"
                  15416. .15674
repeat_region     /rpt_family="ALU"
                  /complement(18751. .19015)
                  /rpt_family="ALU"
                  19128. .19385
repeat_region     /rpt_family="ALU"
                  20516. .20535
repeat_region     /rpt_family="L1"
                  22229. .22478
misc_feature       /note="match to human EST AA018874 (NID:g1482474)
                  ze56g05.s1"
                  /complement(22263. .22510)
                  /note="match to human EST AA058783 (NID:g1551593)
                  zf58f06.r1"
                  /complement(22382. .26553)
                  /gene="WUGSC:GS345D13.2"
                  /complement(join(22382. .22510,26458. .26553))
                  /gene="WUGSC:GS345D13.2"
                  /note="GS345D13.2; match to protein Q08447 (PID:g565181)"

```

| | |
|-----------------------|---|
| | /codon_start=1 |
| | /product="G-protein gamma-1 subunit" |
| | /protein_id="AAB54050.1" |
| | /db_xref="PID:g2078463" |
| | /db_xref="GI:2078463" |
| | /translation="MPVINIEDLTEDKLKMEVDOLKKETVTERMLVSKCEEVRDYY ERSGEDPLVKGIPEDKNPFKEIKGCVIS" |
| misc_feature | complement(22384..22510) |
| | /gene="WUGSC:GS345D13.2" |
| | /note="match to human EST AA015841 (NID:g1476889)" |
| misc_feature | ze30e08.r1" |
| | complement(22388..22888) |
| | /gene="WUGSC:GS345D13.2" |
| | /note="match to human EST W26584 (NID:g1307427)" |
| misc_feature | complement(22394..22510) |
| | /gene="WUGSC:GS345D13.2" |
| | /note="match to human EST AA019183 (NID:g1482573)" |
| misc_feature | ze56g05.r1" |
| | complement(26456..26564) |
| | /note="match to human EST AA015841 (NID:g1476889)" |
| misc_feature | ze30e08.r1" |
| | complement(26456..26564) |
| | /note="match to human EST AA019183 (NID:g1482573)" |
| misc_feature | ze56g05.r1" |
| | complement(26456..26557) |
| Query Match | 11.9%; Score 45.6; DB 10; Length 149764; |
| Best Local Similarity | 49.18; Pred.No.1.7; |
| Matches 11; | Conservative 3; Mismatches 112; Indels 0; Gaps 0; |
| QY 1 | AAACAACAACAAAAAACATCTTCATTCAGAAAATTATCTTAGGACTGATAT 60 |
| | |
| Db 13197 | AAAAAAAAACAAAAACAAAAAGAACAACAAAAACATAATTATTTAACCAATTTAT 13138 |
| QY 61 | TGCTAATTATGGTCAATTTAATWRTTKTGCGGCATTTCCCTTACATGTCTTGACAAGA 120 |
| | : |
| Db 13137 | TAACTTAGGTGCTGAACCTATTTAATTAGGCCACATTCACGTTGATTACTTTCACATAAT 13078 |
| QY 121 | TTAAATGTCTGTGCCAAAATTTGTATTTTATTGAGAGACTTCTTATCMAAAGTAATGC 180 |
| | |
| Db 13077 | TTAAGGTATTTTGACCACACCTTTTAAATTTAAGTAATAGAAAATTTTAAAAATGAATTA 13018 |
| QY 181 | TGCCAAGAAGTCTAAGCAATTAGTAGTGTCCCMTCACCTGTGTT 226 |
| | |
| Db 13017 | GTCACACAAGACAGATATGACATTTGCTAGTCTCCCTGCTTATTT 12972 |
| RESULT 3 | |
| CEY17G7 | |
| LOCUS | CEY17G7 207139 bp DNA HTG 03-DEC-1998 |
| DEFINITION | Caenorhabditis elegans chromosome II clone Y17G7, WORKING DRAFT |
| SEQUENCE | in unordered pieces. |
| Z96049 | |
| ACCESSION | Z96049 |
| NID | 93378012 |
| VERSION | Z96049.1 GI:3378012 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | Caenorhabditis elegans. |
| ORGANISM | Caenorhabditis elegans |
| | Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis. |
| REFERENCE | 1 (bases 1 to 207139) Smye,R. Direct Submission Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or tw@nemastode.wustl.edu On Aug 4, 1998 this sequence version replaced gi:2546913. Order of segments is not known; 800 n's separate segments. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be |
| AUTHORS | |
| TITLE | |
| JOURNAL | |
| COMMENT | |

CDS
join(11683..11753,11836..11928,11980..12304,13242..13462,
14218..14331,14622..14673,18228..18428,18624..18714,
19438..19658)
/gene="Y17G7B.2c"
/note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST
yk470e5.3 comes from this gene; cDNA EST yk470e5.5 comes
from this gene"
/codon_start=1
/protein_id="CA19469.1"
/db_xref="PID:e1356169"
/db_xref="PID:g3947617"
/db_xref="GI:3947617"
/translation="MRSSKGGRGROAPKTAPTTVCCYCDGKRELGSVEVVCSTCLKWF
HGRCLKEFHELSNGVPMICYFTCKQCRPTAEDWKAKADLYOMCVLATLSAER
LKADKLSAEHVPEDEFTYLSLKDEIVPMENWYMLTAIKOKEMHONLAPTLKEKN
IFVQHNDDDLFALEKNLSLGLPLHEAVKLIGKPIERENREPRHIELPIEGPKTR
GASRRHAEPVTKRKQKLAADYSTAAPNGVQIDIPESKDNRYLLEVDPNVPEEP
IFPIFLAFSLFYIEIFQGDVGLIHLVPDKLQIPANLPSEKYLEVSHKGFNLISFK
ANYFEVQESADIAKTLVEMPGSYIEFFHNGKSCGKAYENIAGAYPSISIFKSAT
ATMNLGPKFRNLPRGATGIHARADEQOHEQTLSDMLYLSKEVNLDPKRVKREDDDD
VKDIKEIKOEI"
join(11683..11753,11836..11928,11974..12304,13242..13462,
14218..14331,15865..16127,17187..17299,18228..18428,
18624..18714,19438..19658)
/gene="Y17G7B.2b"
/note="cDNA EST EMBL:D35727 comes from this gene"
/codon_start=1
/protein_id="CA19468.1"
/db_xref="PID:e1323261"
/db_xref="PID:g3947616"
/db_xref="GI:3947616"
/translation="MRSSKGGRGROAPKTAPTTVCCYCDGKRELGSVEVVCSTCLKWF
HGRCLKEFHELSNGVPMICYFTCKQCRPTAEDWKAKADLYOMCVLATLSA
ERLKADKLSAEHVPEDEFTYLSLKDEIVPMENWYMLTAIKOKEMHONLAPTLKE
KNIFVQHNDDDLFALEKNLSLGLPLHEAVKLIGKPIERENREPRHIELPIEGPK
TRGASRRHAEPVTKRKQKLAADYSTAAPNGVQIDIPESKDNRYLLEVDPNVPE
DPAWNOSSAYVIPSFHYRELNPVNVSSNDRAFOLSINGNSITGEGYSMARASH
GVSKGTYFEVNFEDQPDSDSHIRIGMSQSYASLQACVGNKFSYGWSKGTKEHAK
GKKYHGGFKQGDVGLIHLVPDKLQIPANLPSEKYLEVSHKGFNLISFKANYFE
VOESADIAKTLVEMPGSYIEFFHNGKSCGKAYENIAGAYPSISIFKSATATMNLG
PKFRNLPRGATGIHARADEQOHEQTLSDMLYLSKEVNLDPKRVKREDDDDVKDIK
EIKOEI"
20244..23044
/gene="Y17G7B.3"
join(20244..20381,21383..21553,22571..23044)
/gene="Y17G7B.3"
/note="similar to Metallo-beta-lactamase superfamily"
/codon_start=1
/protein_id="CA19450.1"
/db_xref="PID:e1323263"
/db_xref="PID:g3947598"
/db_xref="GI:3947598"
/translation="MYVKSLLRRADNEMYIVKSSSEARALLVDLVNEEDYKELADK
ENIDITAVLTTHHHDGCGNEGFRQFPNVMILGGDSRIPAMDRAHVGMDMAEFAGL
QIKCLSTPCHTSGHICYHITNPADSTSPGVFTGDTLFIAGCGRFEETAPOMDVAL
NEILKNLPVETQIFPGHEVTVANLKFACHVEBGENEKAOKLEMAQRQIEGGFTVPST
VAEEKATNPFMRVRESEEEIQSIGTCDAAVVGMAKLRMKNKE"
24223..29110
/gene="Y17G7B.4"
join(24223..24435,25013..25162,26304..26685,28307..28435,
28869..29110)
/gene="Y17G7B.4"
/note="cDNA EST EMBL:D73584 comes from this gene"
/codon_start=1
/protein_id="CA19451.1"
/db_xref="PID:e1323264"
/db_xref="PID:g3947599"
/db_xref="GI:3947599"
/translation="MSTNEAASQEDIALAQAVLYKSCQVPDGSIPRIGFDESTASC
PDFLSAILSSYSYMGQATHLAQAIQOVNOMLSLDTPLTCDDEKLEPYPEGROKR
SCTIFLGITSNLVTSGLREATHLQVQRNLVDCIVTSAGGIEEDLIKCLKPSYLGFTFM
DGAKLRNGMNRAGNVLIPNDNYCAFEDWMLPIDECLVEOEKHLNWTSPSKLIQRLG
ERIGDESSILYMAKHRIPVFCPALTDGSLGDMLYFHSVKSSPGLRVDIVEDVRHINT

gene
CDS
IAVKSFKTGSIIILGGVVKHHINNALMRNGADHTVYINTGQEFDGSDSGAQPDEAVS
WGKVKPSAGAVKVAEATLVFPLVAETFAKHEGHKD"
complement(32881..37816)
/gene="Y17G7B.5"
complement(join(32881..33253,34402..35645,36629..37493,
37653..37816))
/gene="Y17G7B.5"
/note="similar to MCM2/3/5 family; cDNA EST EMBL:T02370
comes from this gene; cDNA EST yk244e4.5 comes from this
gene"
/codon_start=1
/protein_id="CA19452.1"
/db_xref="PID:e1323265"
/db_xref="PID:g3947600"
/db_xref="GI:3947600"
/translation="MADRANNDVDVROPLPIADDADDVDGIDEMFNNEDEPEDEE
GENLFGDDMERDYREQPELDQYSESGMDASDVGLSVSARRAEREMARQDLODD
ALMYEDGDSEEDTRRRGRGRGRGDAADDSVPMEEEDIPVDILENIRGTIRHVS
DEAVAKEIERERKNFLRSFHEPGNKQTKYIQIKMSAADNRESLEVSFTDLSDNQEO
NISYFLPEAPNEMLAIMDRATEVVMNYPYRSVCNEIKYRISQLPVEEDIRMLROY
HLNMLIRTAGVVTIASGILPOLAVAYKDCVACGYLLGPFVQONDDEVPTICPSCOGK
GPELNVENTYVHYNRYORTMOESPKNVAGRLPRSKDVLILGDCSKPGDEIEYTG
VYTNMFDSLNTKQGFPEVNTLIHANHITNKDKMASDQLTDEDIKAIRLSQDPNISQ
RVFSSIAPSIYGHDDVKRAIALALFREGAKNPGAKHRLRGDINVLGDPGTAKSGFL

Query Match 11.8%; Score 45.2; DB 36; Length 143092;
Best Local Similarity 49.5%; Pred. No. 2;
Matches 140; Conservative 1; Mismatches 140; Indels 2; Gaps 1;
QY 97 TTTCCTTACATGTGCTTGACAGATTAAATGCTGTGCCAAATTTGTGATTATTG 156
DB 17486 TTTCCTTATATTGATTAAAAAAAACAATAAAAAGCTTGAAAAAGTTAGATTAAAG 17545
QY 157 GAGACTTCTTATCAAAAGTATGCTGCCAAAGAGCTTAAGGAATTAGTAGTTC 216
DB 17546 GAGCTTCTGATACCGAATATCAATGCGAAAAAAATCGAAAAATTTCCCTGATTTTATA 17605
QY 217 TCACCTTGTGGAGTGTGCTATCTAAAGATTTTGAATTCCTGGAATGACAATTATATT 276
DB 17606 TTAATTTTAAATCAAAAATTCATTTGTTTAAATTC--AACGACAAATTTTGA 17663
QY 277 TTAACCTTGTGGGGGAANAGTTATAGACACACAGCTTCACTCTGATACCTGTAAT 336
DB 17664 CTGATTACCGTCCGACAGGTGTGTGATTACGAGATTATATTTTAATTCATTTT 17723
QY 337 TAATCTTTAATTCACCTTGTTTGACCATTAAGCTATATGTTT 379
DB 17724 AAAATTTTCATACCAATTTTCTGATTTTCTGTTTGTGTT 17766

RESULT 5
AC006228/c
LOCUS 104308 bp DNA PLN 11-APR-1999
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F5J5, complete
sequence.
AC006228
NID 94580732
VERSION AC006228.4 GI:4580732
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 104308)
AUTHORS Chao,O., Shinn,P., Dunn,P., Buehler,E., Kahn,S., Kim,C., Walker,M.,
Williams,S., Altafi,H., Araujo,R., Conn,L., Conway,A.B.,
Gonzalez,A., Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C.,
Li,J., Liu,S., Lueros,S., Rowley,D., Schwartz,J., Toriumi,M.,
Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and
Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC F5J5

| | |
|--|---|
| JOURNAL | Unpublished (1999) |
| REFERENCE | 2 (bases 1 to 104308) |
| AUTHORS | Ecker,J.R. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (18-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA |
| REFERENCE | 3 (bases 1 to 104308) |
| AUTHORS | Ecker,J.R. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (11-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA On Apr 11, 1999 this sequence version replaced gi:4389503. Location/Qualifiers 1..104308 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" |
| FEATURES | |
| source | |
| BASE COUNT | 33546 a 18958 c 18485 g 33319 t |
| ORIGIN | |
| Query Match | 11.7%; Score 45; DB 8; Length 104308; |
| Best Local Similarity | 48.5%; Pred. No. 2.4; |
| Matches 114; Conservative 3; Mismatches 118; Indels 0; Gaps 0; | |
| OY 19 | AAACAATCTTCATTGAGAAAAAATACTTAGGGACTGATATGGTAATTATGTCAATT 78 |
| Db 56700 | AAATAAATCCTCAAGTTGACACTTATTACAGTGCGTCCACTGCACATTTAATAATTATT 56641 |
| OY 79 | TAAATWRTTKTGGGGCATTCCTTACATGTCTTGACACAGATTAATAATGTCTGCCCAA 138 |
| Db 56640 | TTTGATATATTAGAAGTAATCTTTAAATTAATATTGACATTTAACATCTTCCCAAA 56581 |
| OY 139 | AATTTGTATTTTATTTGGAGACTCTCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAG 198 |
| Db 56580 | GACTCTCTACATTAACTACAGATTAGTTACTAAATAAATTAACCTCCAATAATTAAACAT 56521 |
| OY 199 | GAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTCATAAAGATTTTGA 253 |
| Db 56520 | TATTTAAAAATTACAAAATATATCATTTTGTGATATTGCTTTTTCGACGACTATAA 56466 |
| RESULT 6 | |
| FEMALIP2/c | PFMALIP2 82044 bp DNA HTG 09-APR-1999 |
| LOCUS | Plasmodium falciparum chromosome 1 strain 3D7, WORKING DRAFT |
| DEFINITION | SEQUENCE, in unordered pieces. |
| ACCESSION | AF031745 |
| NID | 94493855 |
| VERSION | AL031745.3 GI:4493855 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | Malaria parasite P. falciparum. |
| ORGANISM | Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. |
| REFERENCE | 1 (bases 1 to 82044) Bowman,S., Churcher,C., Harris,D., Lawson,D., Quail,M. and Barrell,B. |
| AUTHORS | |
| TITLE | Direct Submission |
| JOURNAL | Submitted (09-APR-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK |
| COMMENT | On Mar 24, 1999 this sequence version replaced gi:4455716. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. |

| | | | | | |
|---|--|--|--|--|--|
| | | | | | |
| * NOTE: This is a 'working draft' sequence. | | | | | |
| * This record will be updated with the finished sequence | | | | | |
| * as soon as it is available and the accession number will | | | | | |
| * be preserved. | | | | | |
| Location/Qualifiers | | | | | |
| 1..82044 | | | | | |
| /organism="Plasmodium falciparum" | | | | | |
| /strain="3D7" | | | | | |
| /db_xref="taxon:5833" | | | | | |
| /chromosome="1" | | | | | |
| BASE COUNT 27301 a 7762 c 6747 g 29034 t 11200 others | | | | | |
| ORIGIN | | | | | |
| Query Match 11.6%; Score 44.6; DB 34; Length 82044; | | | | | |
| Best Local Similarity 46.7%; Pred. No. 3; | | | | | |
| Matches 134; Conservative 2; Mismatches 151; Indels 0; Gaps 0; | | | | | |
| QY 1 AAACAACAACAAAAAACAATCTTCATTGAGAAAATTTAGGACTGATAT 60 | | | | | |
| Db 46959 AAAAAAAAAAAAAAAAAATAATATATATATATATATATATTTTATTTATTTATAT 46900 | | | | | |
| QY 61 TGCTAATTATGGTCAATTAATWRTTKTGCGGCATTCCTTACATTGCTTGACAAGA 120 | | | | | |
| Db 46899 ATTACATATATTATTTTTAAATATTATTTTATATATTATTTTATTTTGAATATTT 46840 | | | | | |
| QY 121 TTAATAATGTCTGCCAAAATTTTGATTTTATTTGGAGACTTCTATCAAAGTAATGC 180 | | | | | |
| Db 46839 TTTTATGTATTTTACAAATATAATAATTTTATTCGTATAATAATTATATGCGTATATG 46780 | | | | | |
| QY 181 TGCCAAAGGAAGTCTAAGCAATTAGTAGTGTCCCMTCACCTGTTTGAGAGTGTCTATTC 240 | | | | | |
| Db 46779 TTATATTTCGTGAGTGAGCTAATAAAGTACGTAATAAACAAACGATGATGGAATAT 46720 | | | | | |
| QY 241 TAAAGATTGTGATTTCCCTGGAATGACAATTATTTTAACTTTGGT 287 | | | | | |
| Db 46719 TAATAAAGATGTTACACATGTTAATAATATATATGTTTAAATATGT 46673 | | | | | |
| RESULT 7 | | | | | |
| IPU06960 441 bp DNA INV 18-FEB-1995 | | | | | |
| LOCUS Ichneumon promissorius mitochondrion 16S rRNA gene, partial | | | | | |
| DEFINITION sequence. | | | | | |
| ACCESSION U06960 | | | | | |
| NID 9463865 | | | | | |
| VERSION U06960.1 GI:463865 | | | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| Ichneumon promissorius. | | | | | |
| Mitochondrion Ichneumon promissorius | | | | | |
| Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | | | |
| Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; | | | | | |
| Ichneumonidae; Ichneumoninae; Ichneumon. | | | | | |
| REFERENCE | | | | | |
| AUTHORS Dowton,M. and Austlin,A.D. | | | | | |
| TITLE Molecular phylogeny of the insect order Hymenoptera: apocritan | | | | | |
| JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (21), 9911-9915 (1994) | | | | | |
| MEDLINE 95024071 | | | | | |
| REFERENCE 2 (bases 1 to 441) | | | | | |
| AUTHORS Dowton,M. | | | | | |
| TITLE Direct Submission | | | | | |
| JOURNAL Submitted (21-FEB-1994) Mark Dowton, Department of Crop Protection, | | | | | |
| Weite Campus, University of Adelaide, Adelaide, South Australia | | | | | |
| 5064, Australia | | | | | |
| FEATURES | | | | | |
| source | | | | | |
| 1..441 | | | | | |
| /organism="Ichneumon promissorius" | | | | | |
| /mitochondrion | | | | | |
| /db_xref="taxon:32393" | | | | | |
| /dev_stage="adult" | | | | | |
| <1..>441 | | | | | |
| /note="large subunit ribosomal RNA" | | | | | |
| rRNA | | | | | |

repeat_region 38181. .38243
/note="MER43 repeat: matches 210. .272 of consensus"
repeat_region 38489. .38627
/note="MER43 repeat: matches 99. .236 of consensus"
repeat_region 41519. .42002
/note="MLTID repeat: matches 503. .3 of consensus"
repeat_region 43031. .43115
/note="MIR repeat: matches 71. .155 of consensus"
repeat_region 43180. .43358
/note="MIR repeat: matches 262. .63 of consensus"
repeat_region 43568. .43762
/note="MSTB repeat: matches 1. .195 of consensus"
repeat_region 43763. .44064
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 44119. .44324
/note="AluI repeat: matches 302. .95 of consensus;
incomplete repeat"
repeat_region 46516. .46817
/note="AluYa5 repeat: matches 300. .1 of consensus"
repeat_region 47484. .48286
/note="L1MA5 repeat: matches 1047. .238 of consensus"
repeat_region 48306. .48361
/note="2 copies of 28 mer 93 & conserved"
repeat_region 48401. .48645
/note="L1MA3 repeat: matches 233. .1 of consensus"
repeat_region 48484. .48898
/note="L1 repeat: matches 5390. .4990 of consensus"
repeat_region 48897. .49787
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region 49640. .50086
/note="L1 repeat: matches 5390. .4944 of consensus"
repeat_region 50085. .51156
/note="L1 repeat: matches 3871. .4944 of consensus"
repeat_region 51163. .52381
/note="L1 repeat: matches 4998. .3765 of consensus"
repeat_region 52389. .52714
/note="L1PB2 repeat: matches 899. .550 of consensus"
repeat_region 52705. .54411
/note="L1 repeat: matches 3787. .2073 of consensus"
repeat_region 54704. .54964
/note="L1PA11 repeat: matches 650. .912 of consensus"
repeat_region 56528. .56656
/note="MER5B repeat: matches 1. .130 of consensus"
repeat_region 56528. .56583
/note="MER5A repeat: matches 189. .136 of consensus"
repeat_region 56672. .57300
/note="L1MC3 repeat: matches 2485. .1826 of consensus"
repeat_region 57421. .57473
/note="MER5B repeat: matches 123. .177 of consensus"
repeat_region 59418. .59449
/note="16 copies of 2 mer 84 & conserved"
repeat_region 60640. .60935
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 61816. .61925
/note="MER42C repeat: matches 1421. .1537 of consensus"
repeat_region 63454. .63618
/note="MER20 repeat: matches 45. .218 of consensus"
repeat_region 63715. .64005
/note="AluDb repeat: matches 295. .2 of consensus"
repeat_region 64446. .64604
/note="MLTIB repeat: matches 3. .168 of consensus"
repeat_region 64655. .64774
/note="MLTIA2 repeat: matches 255. .374 of consensus"
repeat_region 65273. .65304
/note="16 copies of 2 mer 91 & conserved"
repeat_region 65305. .66197
/note="L1PA2 repeat: matches 893. .1 of consensus"
repeat_region 66048. .70552
/note="L1 repeat: matches 5390. .887 of consensus"
repeat_region 70539. .71298
/note="L1 repeat: matches 772. .22 of consensus"
repeat_region 74341. .74814
/note="L1 repeat: matches 4921. .5390 of consensus"

repeat_region 74668. .75555
/note="L1PA13 repeat: matches 1. .898 of consensus"
repeat_region 76249. .76352
/note="MIR repeat: matches 191. .82 of consensus"
repeat_region 76424. .76451
Query Match 11.5%; Score 44.2; DB 9; Length 151798;
Best Local Similarity 58.8%; Pred. No. 3;
Matches 70; Conservative 3; Mismatches 46; Indels 0; Gaps 0;
OY 1 AAACAACAACAAAAAACAATCTTCATTCAGAAAAATTACTTAGGACTGATAT 60
Db 26844 AAACAACAACAAAAAATACAAATCTTCTTTATAGTATTTCAAAATCCACCACAT 26903
OY 61 TGGTAATTATGTCATTTAATWRTTTRTGGGCATTTCCCTTACATTTGCTTGACAAG 119
Db 26904 AGCATAATTATTCACAACTGGCTTTGCTGTAGCGCAGTGTGCTGTCATTTCTAGCTAAG 26962
RESULT 9
AC006376/c AC006376 172007 bp DNA PRI 24-MAR-1999
LOCUS Homo sapiens clone NH0386L03, complete sequence.
DEFINITION AC006376
ACCESSION 94508126
NID AC006376.2 GI:4508126
VERSION HTG.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172007)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
TITLE Unpublished
REFERENCE 2 (bases 1 to 172007)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 24, 1999 this sequence version replaced gi:4204339.
FEATURES
source 1. 172007
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0386L03"
BASE COUNT 58408 a 33431 c 31325 g 48843 t
ORIGIN
Query Match 11.5%; Score 44; DB 42; Length 172007;
Best Local Similarity 47.2%; Pred. No. 3.1;
Matches 116; Conservative 5; Mismatches 125; Indels 0; Gaps 0;
OY 40 AATTATCTTAGGAGCATATTTGTAATTAATGTCATTTAATWRTTCTTGGGCAATTT 99
Db 66279 AACTATCTCAGGAATACTGATTAGCACTTACATTTTCAAAATAAATTTGGAATTAC 66220
OY 100 CCTTACATGTCTTGACAAGATTAAATGTCGTGCCAAATTTGTATTTATTTGGAG 159
Db 66219 AGTAATTTTATTTGGTGAAGTGTAGTGCATTTGGTGTATATTTATG 66160
OY 160 ACTTCTTATCAAAAGTAATGCTGCCAAAGAGTCTAAGGAATTAAGTAGTTCCTCCMTC 219
Db 66159 ATAGACATTGAAGGCTAAATAGAAAATGTACTTTATGACACTCTTAATTTAGTGCAAT 66100

[illegible]

```

gene
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="V"
/map="near 60.5 cM"
<1..717
/gene="F21J6.1"
/note="encodes putative P450; entire gene encoded on
T11H3, GenBank accession number AC005964; genomic copy of
EST Z33677; genomic copy of EST Z33693; gene model last
edited on 16 Dec 98"
/evidence=not_experimental
1..24515
/note="overlap with T11H3, GenBank accession number
AC005964, from position 67404 to 91916"
/note="Region: overlap with T11H3"
join(<1..11,103..717)
/gene="F21J6.1"
/note="functional catalog ID=98"
/codon_start=3
/evidence=not_experimental
/product="putative P450"
/protein_id="AAC98443.1"
/db_xref="PID:g4063733"
/db_xref="GI:4063733"
/translation="ILFNILNAGIDTSAQVMTWMTYLLISNPRVKKAAQAEVREVIKNN
KDDIIEDIERLEYLKMVVKETFRVLPVPLIPRESKDYKIGYDIPKKTWIVHNI
WAIHRNPVWKDPEAFIPERENDQIDYKGLNFEFLPGSGRMCPIGGMALVHLT
LINLVRFDWKLPEGMEVEDVDLEESYGLVCPKKVPVLQILPVLTOWS"
complement(1310..1353)
/note="function=putative_microsatellite"
/rpt_type=tandem
/rpt_unit=AT
/4808..6391
/gene="F21J6.2"
/note="encodes putative P450; gene model last edited on 16
Dec 98"
/evidence=not_experimental
join(4808..5683,5777..6391)
/gene="F21J6.2"
/note="functional catalog ID=98"
/codon_start=1
/evidence=not_experimental
/product="putative P450"
/protein_id="AAC98444.1"
/db_xref="PID:g4063734"
/db_xref="GI:4063734"
/translation="MSLWYIIVAFVFESSMIIVIRIKTKNLPGPRLPIGNLHQ
LGSRHSMEKLSSETYGPLMSLKFSGSVSTVASTPEVKEVLTDFDECCSRPMTYP
ARVYNLKDLCFSPYSKYWRKMTVELYTAKRVSQFQTRKEEVALVDFIKQAA
SLEKPVNLKTKMLKLSGSVLCRAFGINLQSKLENTEYEEVIOGTVELVGSFAADYF
PVVGRILDRITGLHSKCEKLRFKAMDAFEDOSIKHLEDEIKDDIIDLKMERGETT
LGEQLTRDHTKGI LANILNAGIDTSAQVMTWMTYLLISNPRVKKAAQAEVREVIKHK
DDIIEDIERLEYLKMVVKETFRINPLVPLIPRESKDYKIGYNI PKKTWIVHNIW
AIHRNPVWKDPEAFIPERENDQIDYKGLNFEFLPGSGRRCPIGGMALVHLTLL
INLVRFDWKLPEGMKVADV DLEESYGLVCPKKIPVLQILPVLTOWT"
23757..24702
/gene="F21J6.6"
/note="mRNA sequence based on a comparison to cDNA L39646"
23757..24702
/gene="F21J6.6"
/note="encodes ZFP3 zinc finger protein; genomic copy of
cDNA L39646"
/evidence=experimental
23757..23810
/gene="F21J6.6"
/note="5'-UTR sequence based on a comparison to cDNA
L39646"
23811..24518
/gene="F21J6.6"
/note="functional catalog ID=98"
/codon_start=1
/evidence=experimental
/product="ZFP3 zinc finger protein"

```



```

/protein_id="AAC98442.1"
/db_xref="PID:94063732"
/db_xref="GI:4063732"
/translation="MDASIVSSSTAAPYQDSLNSIIEDEKDVHNSSHELNLIDCIDD
TTSIVNESTSTEOKLFSCNVCRTFYSSOALGCHONAKRERTLARKGRMAASASA
EGHPYGESPLPEFHGOYNHRSLGQAHSISHKLISYNGFGHYGIINSRLLPEDQPA
IGKFPSMDNFHHHQMMAMAPS VNSRSNNIDS PNTGRVLEGSPTLRWHDKGILL
STSHHEEQOKLDLSKL"
2519.. .24702
/gene="F2IJ6.6"
/note="3'-UTR sequence based on a comparison to CDNA
L39646"
38122.. .38161
/note="function=putative_microsatellite"
/rpt_type=tandem
/rpt_unit=AAT
48069.. .48103
/note="function=putative_microsatellite"
/rpt_type=tandem
/rpt_unit-A
66245.. .66277
/note="function=putative_microsatellite"
/rpt_type=tandem
/rpt_unit=AAC
72338.. .72494
/gene="F2IJ6.18"
/note="5'-UTR sequence based on a comparison to CDNA
X92392"
join(72338.. .73175,73451.. .73571,73660.. .73844,73922.. .74053,
74132.. .74268,74363.. .74677)
/gene="F2IJ6.18"
/note="mRNA sequence based on a comparison to CDNA X92392"
72338.. .74677
/gene="F2IJ6.18"
/note="encodes KNOTTED-like homeobox protein KNOT3;
genomic copy of CDNA X92392; genomic copy of EST T44158;
genomic copy of EST T44387"
/evidence=experimental
join(72495.. .73175,73451.. .73571,73660.. .73844,73922.. .74053,
74132.. .74268,74363.. .74402)
/gene="F2IJ6.18"
/note="functional catalog ID=40.02; functional catalog
ID=50.02"
/codon_start=1
/evidence=experimental
/product="KNAT3 homeodomain protein"
/protein_id="AAC98441.1"
/db_xref="PID:94063731"
/db_xref="GI:4063731"
/translation="MAFHNLSDLSFNHTDQHDPPEPPPPPPOCOOHFOEAPP
PWWLTALLRSSDNNNFLNLHTANTTTASSSDSPSSAAAAAANOWLSRSSSFLQ
RNNNNASIVEGDIDVTGGADMTIOGEMKTGENRGCGATTAGGVSWQNARHK
AEILSHPLEQLSAHVACLRIATPYDQLPRIDAOQAOSGHVAKYALSALGAAGAOLVG
DDKELDFMTHYLVLCSFKEOLOHVRVAMEAVAWCEIESLOSITGVSPGEWG
ATMSDDEQVESDANMFDDGLVDYGFPLIPTESEKLMERVRQELKHHLKOYKEK
IVIDIREILKRRAKGLPGDTTVLKAWWSHWKPYPTTEDRKARLVQETGLQLKQIN
NWFINQRKRWNHSPSSSVTLKNRKS NAGDNSGREFA"
complement(72743.. .72760)
/note="function=trinucleotide_repeat; amino_acid=Ala"
/rpt_type=tandem
/rpt_unit=GCG
74403.. .74677
/gene="F2IJ6.18"
/note="3'-UTR sequence based on a comparison to CDNA
X92392"

Query Match          11.4%; Score 43.8; DB 8; Length 110680;
Best Local Similarity 46.2%; Pred. No. 3.8;
Matches 132; Conservative 3; Mismatches 151; Indels 0; Gaps 0;

CY      18 AAACAATCTTCATTCAGAAAATAATCTTAGGACTGATATTCGTAAATTAATGTCAT 77
        ||| | | | |   | | | | | | | | | | | | | | | | | | | | | |
Db    85243 AAAATTAATCATGCAGTGAGATTATTTACAGCGCGCTGCCACTAGCAATTAATAATTTGT 85302
```

| | | | |
|---------------|---|---|---------------------|
| OY | 78 | TAAATWRTKTKTGGGCAATTCCCTACATGTGCTTGACAAGATTAATAATGCTGTGCCA | 137 |
| Db | 85303 | TTTGTATAATTAGAAGAAAATCTTCTAATTAAATATTGACATTTAACAACTTCGCCAA | 85362 |
| OY | 138 | AAATTTGATTTTAAATTGGAGACTTCTTATCAAAGTAATGCTGCCAAAGGAAGTCTAA | 197 |
| Db | 85363 | AATCTCTACCTTACCCTACACGATTAATTACTAAATAAACTCCAAAAATATTAAATA | 85422 |
| OY | 198 | GGAATTAGTAGTGTCCCMTCACCTGTGTTGGAGTGTGCTATTTCTAAAGATTTGATTTC | 257 |
| Db | 85423 | TTATTTAATTACTTCAAAATATCATTTTTTGATATGCTTTCTACATGATTAATCAT | 85482 |
| OY | 258 | CTGGAATGACAATTATATTTAACCTTTGGTGGGGGGAANAAGTTATA | 303 |
| Db | 85483 | CAATCCGTATAGATATTGATGACATTTAATTACTACAAATTACA | 85528 |
| RESULT | 11 | | |
| LOCUS | HSU209G1 | 36823 bp | DNA PRI 29-OCT-1997 |
| DEFINITION | Human DNA sequence from cosmid U209G1 on chromosome X. | | |
| ACCESSION | Z68873 | | |
| NID | g1164913 | | |
| VERSION | Z68873.1 | GI:1164913 | |
| KEYWORDS | X. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 36823) | | |
| AUTHORS | Odeh,C. | | |
| JOURNAL | Submitted (22-JAN-1996) E-mail enquiries: humquery@sanger.ac.uk | | |
| COMMENT | IMPORTANT: This sequence is the entire insert of clone CU209G1. The true left end of clone U209G1 is at 1 in this sequence. The true right end of clone V1164A6 is at 7813. The true right end of clone U209G1 is at 36823. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX/ . U209G1 is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LL0XNC01. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..36823 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="X" | | |
| | /map="X" | | |
| | /clone_1lb="LL0XNC01" | | |
| | 768..853 | | |
| misc_feature | /note="match: multiple ESTs" | | |
| repeat_region | 1979..2268 | | |
| | /note="Alu repeat: matches 1..308 of consensus" | | |
| repeat_region | 2269..2310 | | |
| | /note="21 copies of 2 mer 81 & conserved" | | |
| repeat_region | 5628..7067 | | |
| | /note="12 copies of 120 mer 86 & conserved" | | |
| repeat_region | 7294..7443 | | |
| | /note="2 copies of 75 mer 82 & conserved" | | |
| repeat_region | 7822..7971 | | |
| | /note="2 copies of 75 mer 86 & conserved" | | |
| repeat_region | 8041..8190 | | |
| | /note="2 copies of 75 mer 83 & conserved" | | |
| repeat_region | 8118..8333 | | |
| | /note="2 copies of 108 mer 81 & conserved" | | |
| repeat_region | 8597..8821 | | |
| | /note="3 copies of 75 mer 81 & conserved" | | |
| misc_feature | 10080..10502 | | |
| | /note="match: 5' EST R79567 clone 145988" | | |
| misc_feature | complement(10976..11105) | | |


```

repeat_region /note="match: 3' EST R79568 clone 145988"
12107. .12148
/note="L1 element fragment"
repeat_region 12233. .12367
/note="L1 element fragment"
repeat_region 12577. .12870
/partial
repeat_region /note="Alu repeat: matches 308. .1 of consensus"
16866. .16892
/note="9 copies of 3 mer 96 & conserved"
repeat_region 17081. .17123
/note="MER41 element fragment"
repeat_region 17583. .17871
/partial
repeat_region /note="Alu repeat: matches 308. .1 of consensus"
18120. .19739
/note="L1 element fragment"
repeat_region 20360. .20484
/note="MSTC element fragment"
repeat_region 21043. .21180
/note="L1 element fragment"
repeat_region 21181. .21474
/partial
repeat_region /note="Alu repeat: matches 308. .1 of consensus"
21528. .21647
/note="L1 element fragment"
repeat_region 21660. .21848
/note="MLT1B element fragment"
repeat_region 21662. .21713
/note="MLT1A element fragment"
repeat_region 21849. .21876
/note="7 copies of 4 mer 96 & conserved; 14 copies of 2
mer 96 & conserved"
repeat_region 21849. .21878
/note="5 copies of 6 mer 93 & conserved"
repeat_region 21927. .22036
/note="MLT1B element fragment"
repeat_region 21966. .22053
/note="MLT1A element fragment"
repeat_region 23206. .23253
/note="MER7 element fragment"
repeat_region 23344. .23570
/note="MER7 element fragment"
repeat_region 23586. .23959
/note="MLT1B element fragment"
repeat_region 23588. .23692
/note="MLT1C element fragment"
repeat_region 23745. .23839
/note="MLT1C element fragment"
repeat_region 24903. .25023
/partial
repeat_region /note="Alu repeat: matches 148. .10 of consensus"
26975. .27080
/partial
repeat_region /note="Alu repeat: matches 306. .193 of consensus"
27432. .27712
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 28071. .28220
/note="5 copies of 30 mer 81 & conserved"
repeat_region 28072. .28203
/note="22 copies of 6 mer 83 & conserved; 33 copies of 4
mer 83 & conserved"
repeat_region 28072. .28201
/note="65 copies of 2 mer 83 & conserved"
repeat_region 28074. .28213
/note="5 copies of 28 mer 81 & conserved"
repeat_region 28266. .28335
/note="MLT2C2 element fragment"
repeat_region 28389. .28628
/note="MLT2C2 element fragment"
repeat_region 28872. .28921
/note="2 copies of 25 mer 96 & conserved"
repeat_region 29191. .29274

```

```

repeat_region      /note="3 copies of 28 mer 82 & conserved"
                   30879. .31173
                   /partial
                   /note="Alu repeat: matches 308. .1 of consensus"
repeat_region      32683. .32857
                   /partial
                   /note="Alu repeat: matches 308. .133 of consensus"
repeat_region      32860. .32906
                   /partial
                   /note="Alu repeat: matches 47. .1 of consensus"
repeat_region      34532. .35401
                   /note="L1 element fragment"
repeat_region      35582. .35988
                   /note="L1 element fragment"
repeat_region      36039. .36219
                   /note="L1 element fragment"
repeat_region      36267. .36563
                   /partial
                   /note="Alu repeat: matches 308. .1 of consensus"
BASE COUNT      10072 a      7172 c      8356 g      11223 t
ORIGIN

```

| | Query Match | 11.48; | Score 43.8; | DB 10; | length 36823; | |
|----|-----------------------|--|----------------|----------------|---------------|---------|
| | Best Local Similarity | 51.08; | Pred. No. 5.1; | | | |
| | Matches 102; | Conservative | 0; | Mismatches 98; | Indels 0; | Gaps 0; |
| QY | 113 | TGACAAGATTAAATGTCTGTGCCAAAATTTTGTATTATTATTTGGAGACTTCTTATCAAA | 172 | | | |
| | | | | | | |
| Db | 31264 | TATTCAGATTAGAGAAGTAATCTTAAACACAGAAATTCACAGTTAAGATCTTATAGA | 31323 | | | |
| QY | 173 | AGTAATGCTGCCAAGGAAGTCTAAGAAATTAGTAGTGTTCCCMTCACCTTGTTGGAGTG | 232 | | | |
| | | | | | | |
| Db | 31324 | AATAATATAAAAAATAAAAAGTACAGAGAAATTAGAACAGCTGCAGTAAGATGATTTAACTT | 31383 | | | |
| QY | 233 | TGCTATTCTAAAGAGATTTTGATTTCCCTGGAATGACAATTATATTTTAACCTTTGGTGGGG | 292 | | | |
| | | | | | | |
| Db | 31384 | TAAATTTTTTAAATTAATGCTGTGTTGTGAGATAACAGTTATATTTTAAATTGAAGATATG | 31443 | | | |
| QY | 293 | AAANAGTTATAGGACCACAG | 312 | | | |
| | | | | | | |
| Db | 31444 | GGCCAGATGATGGACATAG | 31463 | | | |

| RESULT | 12 | | | | |
|------------|--|------------|-----|-------------------------|-------------|
| HS164F3 | | | | | |
| LOCUS | | | | | |
| DEFINITION | HS164F3 | 216991 bp | DNA | HTG | 11-JUN-1999 |
| | Homo sapiens chromosome X clone 164F3, | | | WORKING DRAFT SEQUENCE, | in |
| | unordered pieces. | | | | |
| ACCESSION | AL035422 | | | | |
| NID | g5051831 | | | | |
| VERSION | AL035422.11 | GI:5051831 | | | |
| KEYWORDS | HTG; HTGS_PHASE1. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | | |
| | Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 216991) | | | | |
| AUTHORS | Wilson,S. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humbry@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk On Jun 12, 1999 this sequence version replaced gi:5050941. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dj164F3 Contig_ID: 03649 acc-AL035422 length: 216991 bp. | | | | |
| COMMENT | * NOTE: This is a 'working draft' sequence. | | | | |

Query Match 11.3%; Score 43.2; DB 36; Length 4284;
Best Local Similarity 45.8%; Pred. No. 11;
Matches 135; Conservative 3; Mismatches 157; Indels 0; Gaps 0;

QY 65 AATTATGTCATTAATWRTTKTGGGGCATTCCTTACATGCTCTGACAGATTA 124
||| | | ||||: | | | ||| | | | |
Db 3297 AATTGAAGGATATTAATTTTCTTTTCTTTTCTTATCCCTTTTAAATTC 3356

QY 125 AATGCTGTGCCAAATTTTGTATTTTATTTGAGACTTCTATCAAAAGTATGCTGCC 184
||| | | | | | | | | | | | | | | | |
Db 3357 GGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3416

QY 185 AAAGAGTCTAAGGAATAGTAGTGTCCCTCACTTGTGAGTGTCTATCTTAA 244
| | | | | | | | | | | | | | | | | |
Db 3417 TATGAAATAAATAAATCCAAATAATGGAATACTATTCAAAATAGATTGCTTAT 3476

QY 245 AGATTTGATTTCTCGAATGACATTAATTTTAACCTTGTGGGGAANAGTTATAG 304
||| | | | | | | | | | | | | | | | |
Db 3477 TGATTATTTTCTCGAATTAATAAATAAACAATAAATAAATAAATAAATAATCT 3536

QY 305 GACCACAGCTTCACTTCTGATCTGTAATTAATCTTTTATGCACTGTTT 359
| | | | | | | | | | | | | | | | | |
Db 3537 CCCGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3591

RESULT 14
AC002291 96642 bp DNA PLN 15-JUL-1998
LOCUS AC002291/c
DEFINITION Arabidopsis thaliana chromosome I BAC F22K20 genomic sequence,
complete sequence.
AC002291
NID 92477521
VERSION AC002291.1 GI:2477521
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE 1 (bases 1 to 96642)
AUTHORS Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R.,
Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y.,
Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M.,
Vyotskatala,V.S., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Unpublished (1997)

JOURNAL 2 (bases 1 to 96642)
REFERENCE Federspiel,N.A., Davis,R.W., Conway,A.B., Palm,C.J., Conway,A.R.,
AUTHORS Kurtz,D.B., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Shin,P.,
Sun,H., Oji,O., Osborne,B., Shen,Y.K., Toriumi,M., Vyotskatala,V.,
Yu,G., Theologis,A. and Ecker,J.
Direct Submission
Submitted (05-JUN-1997) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

TITLE 3 (bases 1 to 96642)
JOURNAL Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R.,
AUTHORS Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y.,
Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vyotskatala,V.,
Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (07-OCT-1997) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA

TITLE 4 (bases 1 to 96642)
JOURNAL Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R.,
AUTHORS Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y.,
Oji,O., Osborne,B.I., Shin,P., Sun,H., Toriumi,M., Vyotskatala,V.,
Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (04-FEB-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA

COMMENT 94304, USA
On Oct 7, 1997 this sequence version replaced gi:2251218.
e-mail for correspondence: arabesequence.stanford.edu
Genes with similarity to proteins in the databases are described as
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCAN.html), and NetPlantGene
(S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

FEATURES
source
1. 96642
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F22K20"
1692..3328
/gene="F22K20.1"
join(1692..2051,2148..2235,2316..2448,2542..3328)
/gene="F22K20.1"
/note="similar to 'tub' protein gp|U82468|2072162;
location of EST gb|T45528 and gb|AA395363"
/codon_start=1
/protein_id="AAC00626.1"
/db_xref="pid:g2829918"
/db_xref="GI:2829918"
/translation="MSFRSIVRDVRDSTIGSLSRSFDFKLSLNKEGKSRGVSQDSH
EEOIVTIOETPMANLPELRLDVIKRESESVMPARHVVACASVCRSDMKKEI
VQPELSGKITEPVSLKQGPDPDATMOCFIRKDSNLTHTLYCLSPALVENGKFL
SAKRIRRTYTEYISMADTISRSSNTYIGKIRSNFLGTGKFTYDQPAYNSNIARA
VQPVGLSRFRYSKRVSPKVPSSGYKIAQVSELVNLVGTGRPRMHCAMNSIPASSLAE
GCTVGPQDIIIVRSILDESFRSITSSSRKITDYDSNDFSSARFSDILGPLSEDEV
VLEEGKERNSPPLVKNKPPRWHEQLQCWCLNFRGRTVASVKNQOLIAANOPOPOQ
PQPQPPLTPQPSGQTDGPKIILQFGKVKDMFTMDFRYPLSAFOAFAICLSSFDT
KLACE"
11246..12370
/gene="F22K20.2"
11246..12370
/gene="F22K20.2"
/note="Unknown protein; location of EST gb|AA394987"
/codon_start=1
/protein_id="AAC00632.1"
/db_xref="pid:g2829924"
/db_xref="GI:2829924"
/translation="MESSPVNCLPPDSIHQIFSSLPFRDIMCRSVCKFFNQLTSCQ
FIEIISTRPLNLALRPHHHSRHSNGHATNIRPIYHYDPEQNOFRFNDFL
PERSPOPVASSGDIYLMGDSIDLAESKSLVACNPLTRQFVLPOLGSAMSRHGTVL
VDSVNRVMVTELAALYISGTVANQNLKFSSNLPSPKPSPYLMSSSVFALCDVGSPP
RSQWRLFSCKLTNLTITHTNWCLEKHEMGDIFDIKRPRLRGNGDSKLMIGGLKS
TFSLNPAACSTILILRLDLESLEWEAGRMPLMYRGFOESSKRFVGGGDRVYFSAKR
MGLAMDCWOGWRWIEGVPYADGLCRGFVEDAKLTLP"
11246..15792
/gene="F22K20.3"
/note="Highly similar to atExt1, extensin gene gb|U43627"
complement(14671..15792)
/gene="F22K20.3"
/note="Highly similar to extensin (atExt1) protein
gi|1167961"
/codon_start=1
/product="extensin"
/protein_id="AAC00630.1"
/db_xref="pid:g2829922"
/db_xref="GI:2829922"
/translation="MASFLVLAFLAFVSQTANTANYSSPPPVKHYSPPVYKSPPP
PVKHYSPPVYKSPPPPVKHYSPPVYKSPPVYKHYSPPVYKSPPVYKHYSPPVYK
KHSPPVYKSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYK
SPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYK
PPPVYKSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSPPVYKHYSP

[illegible]

QIVDINHDEDEDSEDSDEFARLAHRSTPTRRPVHGSDSGMVDISGDYKPKQGNSSSQ
GVKKPPPPPHPTSSSSSPVEDASPOQSKSSEVIRNLPppSRHQROQFEHHSS
SGSDSYEGOTRNLSTLSSEPOKEEKPEDLLFKDLVEFAKTRSSKANNNNRSL"
complement(45417) .46193)
/gene="F22K20.8"
complement(45417) .46193)
/gene="F22K20.8"
/note="Hypothetical protein"
/codon_start=1

Query Match 11.2%; Score 42.8; DB 8; Length 96642;
Best Local Similarity 46.0%; Pred. No. 6;
Matches 131; Conservative 3; Mismatches 151; Indels 0; Gaps 0;

gene
CDS

OY 19 AAACAATCTTCATTGAGAAAATTATCTTAGGACTGATATTGGTAATATGTCAATT 78
||| ||| | | | | | | | | | | | | | |
Db 10849 AATAAGTCAATGCAGTTAAGATTATTACAAATGGCTGCACTGACATTTAATAATTATT 10790
||||| ||| | | | | | | | | | | | | | |

OY 79 TAATWRTTRTKTGCGGCATTCCCTTACATGTCTTGACAAGATTAAATGTCTGCCAA 138
| : : | | | | | | | | | | | | | | | |
Db 10789 TTTTATAATTAGAAGAAAAATCTTCTAATTAATATTGATATTTAACAATCTTTCAAA 10730
| | | | | | | | | | | | | | | | | |

OY 139 AATTGTATTTAATTGAGACTCTCTATCAAAGTAATGCTGCCAAGAAGTCTAAG 198
| | | | | | | | | | | | | | | | | |
Db 10729 ATCTTCTACTTTAATACTACACAATTAAATTACTTAAATAAACCCTCCAAAATATTTAAT 10670
| | | | | | | | | | | | | | | | | |

OY 199 GAATTAGTAGTGTCCMTCACTTGTGAGTGTGCTATTTCTAAAGATTGATTGCC 258
| | | | | : | | | | | | | | | | | | | |
Db 10669 CATTTAATTACTACAAAATTATCGTTTTGATATTGCTTTTCGCATGATTATAATCATG 10610
| | | | | | | | | | | | | | | | | |

OY 259 TGGAATGACAATTATATTTTAACCTTGTGGGGGAANAGTTATA 303
| | | | | | | | | | | | | | | | | |
Db 10609 AAGCCGATAGATATTGATAGCATTTAATTACTACAAAATTACA 10565
| | | | | | | | | | | | | | | | | |

RESULT 15
LOCUS ATAC003974/c 92612 bp DNA PLN 19-AUG-1998
DEFINITION Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence,
complete sequence.
ACCESSION AC003974
NID g2914688
VERSION AC003974.1 GI:2914688
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE 1 (bases 1 to 92612)
AUTHORS Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
Somerville,C.R. and Venter,J.C.
Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence
Unpublished
2 (bases 1 to 92612)
AUTHORS Rounsley,S.D. and Lin,X.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 92612)
AUTHORS Rounsley,S.D.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA

e-mail: rounsley@tigr.org
BAC clone F24L7 is from Arabidopsis chromosome II and close to the molecular marker TEn5.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>), and NetPlantgene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at/at.html>).
Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES

source 1. 92612
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="TEn5"
/clone="F24L7"
1. 12476
/note="overlap with BAC clone T21L14 (AC003033:84325..96801)."
misc_feature complement(331..458)
misc_feature /note="exon predicted by xgrail, quality good_shadowexon" 1075..1117
misc_feature /note="exon predicted by xgrail, quality excellent" join(<1184..1650,1741..1906,1992..2042,2132..2381,2566..3919)
gene /gene="F24L7.1" <1184..3919
/gene="F24L7.1" /note="T21L14.21"
CDS join(1184..1650,1741..1906,1992..2042,2132..2381,2566..3713)
/gene="F24L7.1"
/codon_start=1
/product="putative cyclin g-associated kinase"
/protein_id="AAC04479.1"
/db_xref="PID:g2914689"
/db_xref="GI:2914689"
/translation="MWKFKPEAQKEPAGLEGRELEIGNLKQVVRNVAEGSSVYLAQDVNHASKOYALKHMCNDEESLELVMEKISVLKSLGHPNVVTLAHGILDMGRNKK EALLAMDFCGKSLVLENRGAGYEEKQALIFRDVCNAVAFAMHOCSPRIAHRLKA ENLLSSDGQWKLCDFGSVSTNHKIFERAENGIEEDNIRKYYTPPYRAPEMWDLFRR EMISEKVDIMALGCLERICYFNKAFDESKQLQINGNYRIPESPKYSVFITDLIKEM LQASPDERPDIQVKSQFLIVSSQLRKCKMSLTCCSSSATKPSAPRRSPPPPS SGESDSGPLGAFWATQHAKTSVSEDNKNMKEFDEPNSTSKSERVRVDSQPKKPS PVRGARGIQRNKDLETTISQNTTPAANNMTRVSKDADFNSFVADDTTKFDGNK PKKEALAEAIQRLKDELKOTKSEKAETAKFEKLSAICRSQROELDLKOTLASKSA SPSPSRDSSONOPSPGMSHMSSTPSVLFLVYHLSLDTHTKVFIVYLRILIMFLMKOR DKMEGTMWELQODRSNMTSGSDTNSWQFSDPAKPVMEASAKGNNTINQSVTRRSK PASAAGTQGFEPWGFETESFRAAATSAAATSASCTQRMGSGNSTSQRYGNSKMRENQ KTAQPAWGAGF"
misc_feature complement(3972..4026)
/note="exon predicted by xgrail, quality excellent_shadowexon" join(<4325..4879,4961..4993,5123..5175,5266..5305,5403..5480,5590..5697,5791..5870,5967..>6060)

gene /gene="F24L7.2" <4325..>6060
/gene="F24L7.2"
/note="T21L14.22; predicted by genefinder" join(4325..4879,4961..4993,5123..5175,5266..5305,5403..5480,5590..5697,5791..5870,5967..6060)
CDS /gene="F24L7.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC04480.1"
/db_xref="PID:g2914690"
/db_xref="GI:2914690"
/translation="MSTPDNPNPNPNPSVSAVSTPIVTASESPVTQPTNTVTPSSQ PQQTASSYRAIAPLRRHHPQNTYTNPLPIRSNSVTNSPHQPPHPDSSLIYPFG SSGRGFTPIRPVRQNSNVADPVGSPSGYTPRGVYGYHGFVSNLDPMNQFMAA HPQNOQSPQLSGHMKGVPHFLQRLVLSENVVEATPSPTSILDNSGHKAKRSDD ALVLRKRVKRTTEGASLSYSLCRSRLNGAHEGIRIDMTCLPKPLPVDKTEFSLPK DLVEAICEDEKDEESVKHLSSEDLKRRHIDRAKKVRARLREERLKRIRARYKARLAL LLPFGEGCRNE"
misc_feature complement(6589..6697)
/note="exon predicted by xgrail, quality excellent_shadowexon" 6796..6918
misc_feature /note="exon predicted by xgrail, quality marginal" complement(6828..7013)
misc_feature /note="exon predicted by xgrail, quality marginal_shadowexon" 7708..7972
misc_feature /note="exon predicted by xgrail, quality excellent" complement(7752..7844)
misc_feature /note="exon predicted by xgrail, quality excellent_shadowexon" complement(7914..7992)
misc_feature /note="exon predicted by xgrail, quality excellent_shadowexon" complement(8108..8225)
repeat_region /rpt_family="(TAA)n"
misc_feature complement(8448..8593)
/note="exon predicted by xgrail, quality excellent_shadowexon" complement(9131..9157)
repeat_region /rpt_family="AT-rich"
misc_feature complement(9372..9408)
/note="exon predicted by xgrail, quality marginal_shadowexon" join(<9504..9807,9985..10457,10556..>11407)
mRNA /gene="F24L7.3" <9504..>11407
/gene="F24L7.3" /note="T21L14.23"
CDS join(9504..9807,9985..10457,10556..11407)
/gene="F24L7.3"
/codon_start=1
/product="putative phosphate transporter"
/protein_id="AAC04481.1"
/db_xref="PID:g2914691"
/db_xref="GI:2914691"
/translation="MAKKGEVNLNALDAKQTMHYHTAIVIAGMGFFTDAYDLFSISL VTKLGRIRYHVDSKRPGLTPPNVAAVNGVAFCGTLAGLQFEGWLGLGRKKYVG ITIMLMVLCISGSLSGHSANGVMAFLCFRFWLGFGIGDYPLSATIMSEYANKRT RGAFIAAVFAMQGFGLIAGGIVSLVYSTFDHAFKAPTYEDVPSTVPOADYVVRIT IMFGAIPALVLYWRMKMPETARTALVARNTKQAASDMKYLQVDLIAEEAOSNSN SSNPTFTGLTFREFARRHGLHGLTTFWELLDIAYSSNLFQKDIYTAIGWIPAAE TMAIHVEFTVYSKAQTLIALCGTVPGVFTVAFIDILGREFIQLMGFIMFTIEMPALA IPYDHWHRHRENRIGFLIMYSLTMFANFGPNATTFVPAEFPARLRSTCHGISAAG KGAIVGARGFLIYAAOSSDSEKTDAGYPPGIGVRNSLMLACVNFGLIVFTLLVPESK GKSLEISREDEQSGDVTVEMTVANSGRKVPV"
misc_feature 11653..11702
/note="exon predicted by xgrail, quality marginal" complement(12069..12123)
misc_feature /note="exon predicted by xgrail, quality marginal_shadowexon" complement(<12375..>13376)
mRNA


```
gene      /gene="F24L7.4"
          complement(<12375..>13376)
          /gene="F24L7.4"
          /note="predicted by genefinder and genscan"
          complement(12375..13376)
CDS       /gene="F24L7.4"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="AAC04482.1"
          /db_xref="PID:g2914692"
          /db_xref="GI:2914692"
          /translation="MVMKIAIATERDLPRLVSTAIRASNSQNASDVERCFDYLRLK
          GLNLSVKNLSHSHKVIPLPESLRDHPNKIRTEAHVLFISWMTFYSSGQNSSSTCNKP
          NPLKIKKVVYKACSELKKNEQLSHGFAYLKAKETVFLGKKNEDESRVHETREMK
          QIGDSKSFALMRTIEKKQOSAPIOGSKNPRSGAGETGVILKILRKPDPFCCSVPSRP
          PLNMKKHQPVPKAFENPKTCLVLKKNSEMLEFEMAKKSADVANAAGFLAAKEASIC
          VDILALMRFSIISTAIETRRIMEKLERLTKHKDRKICNALALLHWRQTI RNQQ"
          repeat_region
          complement(13364..13410)

Query Match      11.28; Score 42.8; DB 8; Length 92612;
Best Local Similarity 46.0%; Pred. No. 6.1;
Matches 131; Conservative 3; Mismatches 151; Indels 0; Gaps 0;

QY  19  AACAAATCTTCATTCAGAAAATATCTTAGGAGCTGATATTGTAATTAGGTCATT 78
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  53198 AAATAAATCCTGGAGTTGAGATTATTACAGTCGCTACCACTGCATTATAATTGTT 53139

QY  79  TAAATWRTTRTTGGGCAATTCCTTACATTGCTTGACAAGATTAAAGTCTGTGCCAA 138
      | : || | | | | | | | | | | | | | | | | | | | | | | |
Db  53138 TTGATTAATTAGAAAGAAATCTTAATTAAATATTGACATTTAACAATCTTCCCAA 53079

QY  139 AATTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAAGTCTAAG 198
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  53078 ATCTTCTACCTTAACACAAATTAATTAATAAATAAACTTCCAAAATATTTAATAT 53019

QY  199 GAATTAGTAGTGTCCCMTCACCTGTTTGAGAGTGTGCTATTCTAAAAGATTGATTCC 258
      | | | | | : | | | | | | | | | | | | | | | | | | | |
Db  53018 CATTTAATTACTACAAAATTAATCAATTTTGTGATATGCTTTTCGACATGATTAATCATC 52959

QY  259 TGGAAATGACAATTAATTTAACTTTGGTGGGGGAANAAGTTATA 303
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  52958 AAGCCGATAGATATTGATAGCAATTAATTACTACAAAATTACA 52914
```

Search completed: September 25, 1999, 23:44:03
Job time: 7786 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:57:21 ; Search time 273 Seconds
(without alignments)
351.002 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383

Sequence: 1 AAACCAACCAACAAAAA.....ATTAGCTATGTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| 1 | 380 | 99.2 | 383 | 1 | V58693 | Prostate tumour sp |
| 2 | 380 | 99.2 | 383 | 1 | V61268 | CDNA sequence of p |
| 3 | 40 | 10.4 | 110000 | 1 | X20248_08 | Continuation (9 of |
| 4 | 39.8 | 10.4 | 110000 | 1 | X21209_03 | Continuation (4 of |
| 5 | 39.4 | 10.3 | 110000 | 1 | X20248_06 | Continuation (7 of |
| 6 | 39.4 | 10.3 | 111309 | 1 | X20250 | Borrelia burgdorfer |
| 7 | 38.6 | 10.1 | 2287 | 1 | V22738 | Babesia microti BM |
| 8 | 38 | 9.9 | 4093 | 1 | V74419 | Staphylococcus aur |
| 9 | 37.6 | 9.8 | 2636 | 1 | V01871 | Human brain specif |
| 10 | 37.2 | 9.7 | 685 | 1 | X19427 | Nicotiana tabacum |
| 11 | 36.8 | 9.6 | 110000 | 1 | V21209_07 | Continuation (8 of |
| 12 | 36 | 9.4 | 2418 | 1 | Q27886 | P. falciparum GBP13 |
| 13 | 36 | 9.4 | 3607 | 1 | T89990 | Cryptosporidium pa |
| 14 | 35.8 | 9.3 | 110000 | 1 | V21209_05 | Continuation (6 of |
| 15 | 35.6 | 9.3 | 9047 | 1 | X12949 | Enterococcus faeca |
| 16 | 35.2 | 9.2 | 731 | 1 | Q21036 | Region specific to |
| 17 | 35.2 | 9.2 | 5064 | 1 | V69110 | Neisseria meningit |
| 18 | 35.2 | 9.2 | 400 | 1 | V69123 | Neisseria meningit |
| 19 | 35 | 9.1 | 8371 | 1 | T66909 | Human neuropeptide |
| 20 | 35 | 9.1 | 110000 | 1 | V21209_00 | Methanococcus jann |
| 21 | 35 | 9.1 | 600 | 1 | V86185 | EST clone J635. Ne |
| 22 | 35 | 9.1 | 14752 | 1 | X20256 | Borrelia burgdorfer |
| 23 | 34.8 | 9.1 | 10182 | 1 | V74427 | Staphylococcus aur |
| 24 | 34.6 | 9.0 | 5534 | 1 | Q35988 | Tomato hsp80 genom |
| 25 | 34.6 | 9.0 | 2334 | 1 | V09695 | M. hyopneumoniae 2 |
| 26 | 34.6 | 9.0 | 18359 | 1 | X20255 | Borrelia burgdorfer |
| 27 | 34.6 | 9.0 | 9842 | 1 | X20259 | Borrelia burgdorfer |
| 28 | 34.4 | 9.0 | 110000 | 1 | T58840_4 | Continuation (5 of |
| 29 | 34.4 | 9.0 | 3198 | 1 | T92702 | Candida Carhol gen |
| 30 | 34.4 | 9.0 | 3198 | 1 | T92869 | Candida Carhol gen |
| 31 | 34.4 | 9.0 | 1158 | 1 | V24720 | H. pylori ORF 06ep |
| 32 | 34.4 | 9.0 | 1146 | 1 | V25046 | H. pylori cytoplasm |
| 33 | 34.4 | 9.0 | 1350 | 1 | X14148 | H. pylori GHPD 422 |
| 34 | 34.4 | 9.0 | 110000 | 1 | X20248_05 | Continuation (6 of |
| 35 | 34.4 | 9.0 | 110000 | 1 | X20248_05 | Continuation (6 of |
| 36 | 34.4 | 9.0 | 116277 | 1 | X20249 | Borrelia burgdorfer |
| 37 | 34.2 | 8.9 | 2749 | 1 | T66652 | Housefly Hermes el |
| 38 | 34.2 | 8.9 | 3164 | 1 | V26159 | Swinepox virus Hin |
| 39 | 34.2 | 8.9 | 1762 | 1 | V33193 | Secreted protein C |
| 40 | 34.2 | 8.9 | 14752 | 1 | X20256 | Borrelia burgdorfer |
| 41 | 34 | 8.9 | 5852 | 1 | Q11710 | Dictyostelium plas |
| 42 | 34 | 8.9 | 2503 | 1 | O53480 | PNPX30 xylanase CD |
| 43 | 34 | 8.9 | 19124 | 1 | T72882 | Plasmodium var-7 g |

ALIGNMENTS

| | | | | | | | |
|---|----|----|-----|-------|---|--------|--------------------|
| C | 44 | 34 | 8.9 | 5849 | 1 | V33135 | Plasmodium berghel |
| C | 45 | 34 | 8.9 | 11802 | 1 | V74381 | Staphylococcus aur |

ALIGNMENTS

| | |
|----------|--|
| RESULT | 1 |
| V58693 | |
| ID | V58693 standard; cDNA; 383 BP. |
| AC | V58693; |
| DR | 08-DEC-1998 (first entry) |
| DE | Prostate tumour specific gene clone P509S. |
| KW | Prostate tumour specific gene; human; prostate cancer; detection; |
| KW | therapy; ss. |
| OS | Homo sapiens. |
| PN | WO9837418-A2. |
| PD | 27-AUG-1998. |
| PF | 25-FEB-1998; U03690. |
| PR | 09-FEB-1998; US-904809. |
| PR | 25-FEB-1997; US-806596. |
| PR | 01-AUG-1997; US-904809. |
| PA | (CORI-) CORIXA CORP. |
| PI | Dillon DC, Xu J; |
| DR | WPI; 98-480805/41. |
| PT | Novel human prostate specific tumour protein and fragments - useful |
| PT | for detecting and treating prostate cancers |
| PS | Claim 1; Page 134; 141pp; English. |
| CC | This sequence represents a human prostate tumour specific gene, and can |
| CC | be used in the method of the invention. The method is for detecting |
| CC | prostate cancer comprises contacting a biological sample with an agent |
| CC | able to bind an immunogenic portion of a prostate protein (such as |
| CC | encoded by this sequence). An antibody which binds to an immunogenic |
| CC | portion of the prostate protein, and the method can be used to detect, |
| CC | monitor progression of, or treat prostate cancers. The antibody may |
| CC | also be conjugated to a therapeutic agent for use in therapy of prostate |
| CC | cancers. |
| Sequence | 383 BP; 123 A; 51 C; 63 G; 140 T; |

| Query Match | 99.28; | Score 380; | DB 1; | Length 383; |
|-----------------------|---|--------------------|------------|-------------|
| Best Local Similarity | 100.04; | Pred. No. 7.3e-77; | | |
| Matches 383; | Conservative | 0; | Mismatches | 0; |
| | | | Indels | 0; |
| | | | Gaps | 0; |
| QY 1 | AAACACAACAACAAAAAACAATCTCTCATTCAGAAAAATATCTTAGGCACTGATAT | 60 | | |
| DB 1 | AAACACAACAACAAAAAACAATCTCTCATTCAGAAAAATATCTTAGGCACTGATAT | 60 | | |
| QY 61 | TGTAATATATGTCATTTAAATWTRRTKTGGGGCAATTCCTTACATTTGCTTGACAAGA | 120 | | |
| DB 61 | TGTAATATATGTCATTTAAATWTRRTKTGGGGCAATTCCTTACATTTGCTTGACAAGA | 120 | | |
| QY 121 | TTAAATGCTCTGTGCCAAAAATTTGTATTTTATTTGGAGACTTCTTATCAAAAGTAATGC | 180 | | |
| DB 121 | TTAAATGCTCTGTGCCAAAAATTTGTATTTTATTTGGAGACTTCTTATCAAAAGTAATGC | 180 | | |
| QY 181 | TGCCAAGAAGAGCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC | 240 | | |
| DB 181 | TGCCAAGAAGAGCTAAGGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC | 240 | | |
| QY 241 | TAAAGATTTTGTATTTCCCTGGAATGACAAATTAATTTTAACCTTTGGTGGGGGAANAGTT | 300 | | |
| DB 241 | TAAAGATTTTGTATTTCCCTGGAATGACAAATTAATTTTAACCTTTGGTGGGGGAANAGTT | 300 | | |
| QY 301 | ATAGACACACAGTCTTCACCTCTGATACTGTAAATTAATCTTTATTTGCACCTGTTTG | 360 | | |
| DB 301 | ATAGACACACAGTCTTCACCTCTGATACTGTAAATTAATCTTTATTTGCACCTGTTTG | 360 | | |
| QY 361 | ACCATTAAGCTATATGTTTAAA | 383 | | |
| DB 361 | ACCATTAAGCTATATGTTTAAA | 383 | | |

RESULT 2
ID V61268 standard; cDNA; 383 BP.
AC V61268;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone p509S.
KW prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-AZ.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS Claim 3; Page 124; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
of a prostate tumour protein. The encoded immunogen, or the DNA itself,
can be used as a vaccine for the treatment of prostate cancer. The DNA
was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 383 BP; 123 A; 51 C; 63 G; 140 T;

Query Match 99.2%; Score 380; DB 1; Length 383;
Best Local Similarity 100.0%; Pred. No. 7.3e-77;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAACAACAACAAAAACAATCTCTCATTCAGAAAAATATATCTTAGGACTGATAT 60
DB 1 AAAACAACAACAAAAACAATCTCTCATTCAGAAAAATATATCTTAGGACTGATAT 60
QY 61 TGGTAATATGTCATTTATWTRTKTGTGGGCAATTCCTTACATTGCTTGACAAGA 120
DB 61 TGGTAATATGTCATTTATWTRTKTGTGGGCAATTCCTTACATTGCTTGACAAGA 120
QY 121 TTTAAATGCTGTGCCAAAAATTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGC 180
DB 121 TTTAAATGCTGTGCCAAAAATTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGC 180
QY 181 TGGCAAGGAAGTCTAAGAAATAGTAGTGTCCCMTCACCTGTTTGAGAGTGTCTATTC 240
DB 181 TGGCAAGGAAGTCTAAGAAATAGTAGTGTCCCMTCACCTGTTTGAGAGTGTCTATTC 240
QY 241 TAAAGATTTTGTATTCCTGGAATGACAATATATTTAACTTTGTGGGGGAANAGTT 300
DB 241 TAAAGATTTTGTATTCCTGGAATGACAATATATTTAACTTTGTGGGGGAANAGTT 300
QY 301 ATAGACACACAGTCTTCACCTTCTGACTGTGAATTAATCTTTATTCGACTGTTTG 360
DB 301 ATAGACACACAGTCTTCACCTTCTGACTGTGAATTAATCTTTATTCGACTGTTTG 360
QY 361 ACCATTAGCTATATGTTTAAA 383
DB 361 ACCATTAGCTATATGTTTAAA 383

RESULT 3
X20248_08
Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide s
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000

WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 10.4%; Score 40; DB 1; Length 110000;
Best Local Similarity 46.5%; Pred. No. 1.1;
Matches 155; Conservative 4; Mismatches 165; Indels 9; Gaps 1;

QY 38 AAAATTATCTTAGGACTGATATTTGTAATTAATGTCATTTATWTRTKTGGGCAT 97
DB 14020 AAGAGTTTCTAATGAAAAATTTGAGAGATATATAGTGTTTTATGATGAGGTTTAAGAAT 14079
QY 98 TTCCTTACATTGCTTGACAAGATTAAAAATGCTGTGCCAAAAATTTGTATTATTG 157
DB 14080 AATTTATTTTGGCAATAAATAGTATATCAACATCTTAATGTTCAAGATTTTATCTGA 14139
QY 158 AGACTTCTTATCAAA-----AGTAATGCTGCCAAAGAGTCTAAGAAATTAGTAG 208
DB 14140 TTTATTTATTTTAAACCAAGAGCTAAAAAATTTGCTTAATGAATGAAGAATATAGA 14199
QY 209 TGTCCCMTCACCTGTTTGAGAGTGTGCTATCTTAAAGATTGATTTCTCGAATGACA 268
DB 14200 TGTTCATGCTCTGTAATGAAAGTGGTGTGTTAAAGAAATATGATTCGATTGAATTGA 14259
QY 269 ATTATATTTAACCTTGTGGGGGAANAGTTATAGGACCACAGTCTTCACTCTGTATAC 328
DB 14260 AATTAATAATTTAATAGACGAGTTGAAAAATCAGATAAAAAATAATAGTTGAACGC 14319
QY 329 TTGTAATTAATCTTTATTTGACCTGTTTGA 361
DB 14320 AGAATATTGACTTCTAATTTTGTGTTGTAATA 14352

RESULT 4
V21209_03/c
Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match 10.4%; Score 39.8; DB 1; Length 110000;
Best Local Similarity 47.9%; Pred. No. 1.2;
Matches 101; Conservative 4; Mismatches 106; Indels 0; Gaps 0;

QY 1 AAAACAACAACAAAAACAATCTCTCATTCAGAAAAATATCTTAGGACTGATAT 60
DB 41592 AAGATCCAGAGACAAAGACTGATGTTAGAGTGAAGATACATTAATGGAGAGATTG 41533
QY 61 TGGTAATATGTCATTTATWTRTKTGGGCAATTCCTTACATTGCTTGACAAGA 120
DB 41532 TAGAAATGAGCAGATTTAGTGTATTTGTCAGCAGATTGTACCAAGACGACAATC 41473
QY 121 TTTAAATGCTGTGCCAAAAATTTGTATTATTGAGAGACTTCTTATCAAAAGTAATGC 180

DB 41472 CAAATTTGGCTAGATGCTTGGTTAGAGCTCAGTCCAGATGCAATCTTCAAGAGTTC 41413
OY 181 TGCCAAGAGAGTCTAAGCAATAGTAGTCT 211
DB 41412 ATCCAAGAGTACCTCCAGTTAATACAAAGGT 41382

RESULT 5

X20248_06
Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Sequence Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 10.3%; Score 39.4; DB 1; Length 110000;
Best Local Similarity 43.5%; Pred. No. 1.5;

Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

OY 29 TCATTCAGAAAAATATCTTAGGAGCTGATTTGGTAATATGTCATTTAATWTRRTT 88
DB 76486 TGATTCCTTAACCTTATTTGTTAAAGACTTAAAGCTAATGTTAAATGAATTTAAAGTCTT 76545
OY 89 KTGCGCATTTCCCTTACATGTCTGACACAGATTAATGCTGTCGCCAAATTTTGTAT 148
DB 76546 TCCATCGATATCCCTGCTGAGCTTTCAGATATATTATTGCCCCAAAGTATTGGAT 76605
OY 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAG 208
DB 76606 TTTATTTGTTAATGTTTATGTAAGAATAATTTTGTAGATAAAATCTTGAAGAGCTCTTTT 76665
OY 209 TGTTCCTCCTCACTTGTGAGTGTGCTATTTCTAAAGATTTGATTTCTCGAATGACA 268
DB 76666 GAATAAAGATAGGTATATGTAATCTTTCTTCTAAAAAATCTAATTTAAAAAATAATT 76725
OY 269 ATTATATTTTAACTTTGTTGGGGGAANAGTTATAGGACACAGCTTCTCACTTCGATAC 328
DB 76726 TACATTTGAAATATTTGTTACTCTTAACCTTAAAGATGTAGATTAGCTCTTTTATTC 76785
OY 329 TTGTAATTAATCTTTTATTTGACACTTTTGTGACCATTAAGCTATATGTTTAA 382
DB 76786 ATTAAGATCCAAACTTTTCTTATTTGTTTGGCAAACTCTTCTATTTTATGTTAA 76839

RESULT 6

X20250
ID X20250 standard; DNA; 111309 BP.
AC X20250;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #3.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMAN-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,

PI White OR;
DR WPI; 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 738-800; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T;

Query Match 10.3%; Score 39.4; DB 1; Length 111309;
Best Local Similarity 43.5%; Pred. No. 1.5;

Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

OY 29 TCATTCAGAAAAATATCTTAGGAGCTGATTTGGTAATATGTCATTTAATWTRRTT 88
DB 87100 TGATTCCTTAACCTTATTTGTTAAAGACTTAAAGCTAATGTTAAATGAATTTAAAGTCTT 87159
OY 89 KTGCGCATTTCCCTTACATGTCTGACACAGATTAATGCTGTGCCAAATTTTGTAT 148
DB 87160 TCCATCGATATCCCTGCTGAGCTTTCAGATATATTATTGCCCCAAAGTATTGGAT 87219
OY 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAG 208
DB 87220 TTTATTTGTTAATGTTTATGTAAGAATAATTTTGTAGATAAAATCTTGAAGAGCTCTTT 87279
OY 209 TGTTCCTCCTCACTTGTGAGTGTGCTATTTCTAAAGATTTGATTTCTCGAATGACA 268
DB 87280 GAATAAAGATAGGTATATGTAATCTTTCTTCTAAAAAATCTAATTTAAAAAATAATT 87339
OY 269 ATTATATTTTAACTTTGTTGGGGGAANAGTTATAGGACACAGCTTCTCACTTCGATAC 328
DB 87340 TACATTTGAAATATTTGTTACTCTTAACCTTAAAGATGTAGATTAGCTCTTTATTC 87399
OY 329 TTGTAATTAATCTTTTATTTGACACTTTTGTGACCATTAAGCTATATGTTTAA 382
DB 87400 ATTAAGATCCAAACTTTTCTTATTTGTTTGGCAAACTCTTCTATTTTATGTTAA 87453

RESULT 7

V22738/c
ID V22738 standard; DNA; 2287 BP.
AC V22738;
DT 28-SEP-1998 (first entry)
DE Babesia microti BMN1-8 antigen sequence.
KW antigen; detection; diagnosis; vaccine; tick-borne disease;
KW differentiation; Lyme disease; ehrlichiosis; ss.
OS Babesia microti.
FT key location/Qualifiers
FT CDS 31.1806
FT /*tag= a
FT /product= antigen
PN EP-834567-A2.
PD 08-APR-1998.
PF 01-OCT-1997; 117067.
PR 24-APR-1997; US-845258.
PR 01-OCT-1996; US-723142.
PI (CORI-) CORIXA CORP.
PI Houghton R, Lodes MJ, Reed SG, Sleath PR;
DR WPI; 98-195465/18.
DR P-PSDB; W56287.
PT Polypeptides comprising Babesia microti antigens and their
PT immunogenic fragments or epitopes - and related nucleic acid,
PT vectors, transformed cells and antibodies, useful for diagnosis of
PT infection and in protective vaccines

PA (SARA) OTSUKA PHARM CO LTD.
PI Fujiwara T, Horie M, Watanabe T;
DR WPI: 97-459830/43.
DR P-PSDB; W37495.
PT Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT useful for diagnosis or therapy of hereditary disease and cancer
PS Claim 6; Page 60-64; 123pp; English.
CC The present sequence encodes a brain specific nucleosome assembly
CC protein isolated from a human foetal brain cDNA library. The nucleotide
CC or amino acid sequences are useful for in-vitro diagnosis of hereditary
CC diseases and cancer and for preparation of pharmaceuticals.
SQ Sequence 2636 BP; 814 A; 521 C; 599 G; 702 T;

Query Match 9.8%; Score 37.6; DB 1; Length 2636;
Best Local Similarity 46.6%; Pred. No. 2.5;
Matches 109; Conservative 3; Mismatches 122; Indels 0; Gaps 0;

OY 55 TGATATTGTAATTAAGTCATTTAATWRTTKTGGGCATTTCCCTTACATGTCTTG 114
DB 2039 TGCTTTAGATAATGACACAGCTTACATATTTGAAGAATATTACTACACAGCCTA 1980
OY 115 ACAAGATTAATGTCGTGCCAAATTTGTAATTTATTTGGAGACTTCTATCAAAAG 174
DB 1979 GATAGATAAAGTGGCACTTAGACACTTTTGAAGACTATTTTAAATATAGACTACCAA 1920
OY 175 TAATGCTGCCAAAGAGTCTAAGGAATTAGTGTCCCMTCACCTGTTGGAGTGTG 234
DB 1919 GAAACCACAAAACTGACTATAGAATAATAGTGTTCAGGTTACAGGTCAAGGTTTAA 1860
OY 235 CTATTTCTAAAGATTTTGATTTCTCTGGAATGACAATTAATTTTAACTTGGTG 288
DB 1859 CTTTTTCAAGGCTGTATGATCTGCTTCACTTCTTGAGATTTTAAGATTTCTTG 1806

RESULT 10
X19427
ID X19427 standard; DNA; 685 BP.
AC X19427;
DT 04-JUN-1999 (first entry)
DE Nicotiana tabacum matrix attachment region ps211-1 DNA sequence.
KW Nicotiana tabacum; matrix attachment region; MAR; tobacco;
KW expression; transgene; structural gene; ss.
OS Nicotiana tabacum.
PN WO9907866-A1.
PD 18-FEB-1999.
PF 05-AUG-1998; U16344.
PR 06-AUG-1997; US-066118.
PA (UNNC-) UNIV NORTH CAROLINA STATE.
PI Michalowski SM, Spiker SL;
DR WPI: 99-167436/14.
PT New Matrix Attachment Region DNA - useful for increasing expression
PT of transgene produce, especially structural genes in tobacco
PS Claim 1; Fig 3; 66pp; English.
CC The present sequence represents DNA encoding a Matrix Attachment Region
CC (MAR) (which bind to the nuclear matrix) isolated from Nicotiana tabacum.
CC A DNA construct containing MAR sequences can be used to increase
CC expression of a transgene product. A new method identifies MARS by
CC sequence alone, unlike prior art methods using associated genes. MARS
CC increase expression of a transgene product.
SQ Sequence 685 BP; 243 A; 107 C; 87 G; 248 T;

Query Match 9.7%; Score 37.2; DB 1; Length 685;
Best Local Similarity 49.4%; Pred. No. 2.6;
Matches 87; Conservative 3; Mismatches 86; Indels 0; Gaps 0;

OY 31 ATTCAGAAAAATATCTTAGGAGCTGATATGTAATATGTCATTTAATWRTTKT 90
DB 366 ATTAAAAAATAAGATGGGGCCCTGCATTTTATTTAATATTTAATTTT 425
OY 91 GGGGCAATTCCTTACATGTCCTTGACAAGATTAATAATGCTGTGCCAAATTTTGTATT 150

DB 426 AGCGAAGATCCCTCCCTTATTTATGAAATACCCCTTAATGACTACATCTTATTAAT 485
OY 151 TATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGCTTAAGAAATAGT 206
DB 486 AAGTTGTCTATATATATGAAGTCAATCTCTACATACATTAATAACATATTAA 541

RESULT 11
V21209_07
Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

Query Match 9.6%; Score 36.8; DB 1; Length 110000;
Best Local Similarity 45.2%; Pred. No. 5.7;
Matches 122; Conservative 3; Mismatches 145; Indels 0; Gaps 0;

OY 18 AAACAATTCCTTCATTCAGAAAAATATCTTAGGAGCTGATATGTAATTAATGTCAT 77
DB 64033 ATAAATATATTCACACTTTTATATTTATTTTGCATTAAATTAATTAATTTTAA 64092
OY 78 TTAATWRTTKTGGGCAATTTCCCTTACATGTCCTTGACAAGATTAATAATGCTGTGCA 137
DB 64093 ATTTTCTGACTTAAATATTTTAAATGTAATTTTTCGCAAAAAAAGAACAAATTTCC 64152
OY 138 AAATTTGATTTATTTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGCTTAA 197
DB 64153 AAATTTTAAATGATTTTATAGAAAAAATGAAAAATTAATTTTAAATAAAAAAT 64212
OY 198 GGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATCTTAAAGATTTTGATTTC 257
DB 64213 TAGAAATTAAGTTAAATTAACAATTTTGAATTTAAATTTTACCTATCCCTTAATTA 64272
OY 258 CTGAATGACAATTAATTTTAACCTTGGT 287
DB 64273 TCCTTTGAGACATTAATTTTACATTTTGT 64302

RESULT 12
O27886/c
ID O27886 standard; DNA; 2418 BP.
AC O27886;
DT 04-FEB-1993 (first entry)
DE P.falciparum GBP130h.
KW Polymerase chain reaction; glycophorin binding protein; inverse PCR;
KW Plasmodium falciparum; malaria; ss.
OS Synthetic.
FH Key
FT exon
FT exon
FT exon
FT exon
FT exon
FT exon
FT exon
FT repeat_region

FT EP-499834-A. /tag= c
PN 26-AUG-1992.
PD 27-JAN-1992; 101271.
PF 21-FEB-1991; DE-105348.
PR (BEHM) BEHRINGWERKE AG.
PI Hundt E, Knapp B, Kupper H, Nolte D, Kuepper H;
DR WPI; 92-286009/35.
PT DNA coding protein GBP 130 h and proteins produced - used for
PS vaccines to control malaria
CC Claim 1; Page 12-14; 17pp; German.
CC Glycophorin binding protein 130h is a P.falciparum blood stage
CC antigen 698 homologous with the known GBP130. The GBP130h gene and
CC its homologues can be used in vaccines to protect against malaria
CC caused by P.falciparum. The coding sequence was isolated using PCR
CC techniques on a genomic P.falciparum FCBR strain library. The
CC GBP130h gene was found to be highly conserved between different
CC strains but is distinct from the GBP130 gene.
CC See Q27878-Q27885 and Q33432.
SQ Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T;

Query Match 9.4%; Score 36; DB 1; Length 2418;
Best Local Similarity 42.6%; Pred. No. 5.6;
Matches 159; Conservative 4; Mismatches 210; Indels 0; Gaps 0;

OY 1 AAAACAACAACAAAAACAATTCATCTCAGAAAAATATCTTAGGACTGATAT 60
DB 638 ATAAAAATAAAAAAACAATACATAATATTTTATTTATTTATTTATTTT 579
OY 61 TGGTAATATGTCATTTAATWRTTKTGGGCATTCCTTACATTTGCTTGACAAGA 120
DB 578 TTTTTTTACATAAATTACATTTCAATGTTTTCATATAAAGATAAATACTATATAT 519
OY 121 TTAAGATGCTGTGCCAAATTTGTATTTTATTTGGAGACTCTTATCAAAAGTAATGC 180
DB 518 ATATTTTTCATGATATATAATATATCATATCCTTGAAGAAATATATGATATATTTAAT 459
OY 181 TGCCAAAGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTC 240
DB 458 TAGTAAGTTACAGAAGCTAGAATAATATATGTATGACTTACGACTATGATATATA 399
OY 241 TAAAGATTTGATTTCCGTGAATGACAATATATTTAACTTTGGTGGGAANAGTT 300
DB 398 AAATAATTAATTTATTCATGATATATTTATTTTGAATCTTAAATTAATAAAGAA 339
OY 301 ATAGACACAGTCTCTCACTCTGACTTGTAAATTAATCTTTTATGCACTGTTTG 360
DB 338 TTTAGATAATATTTTATATATATATTTTCCACATGTTCTTTTGTATCTATCTTTTC 279
OY 361 ACCATTAAGCTAT 373
DB 278 TATTAATAAATAT 266

RESULT 13
T89990/c
ID T89990 standard; DNA; 3607 BP.
AC T89990;
DT 11-MAY-1998 (first entry)
DE Cryptosporidium parvum oocyst heat shock protein HSP70 gene.
KW Cryptosporidium; detection; assay; protozoan; heat shock protein;
KW HSP70; gene amplification; PCR; ds.
OS Cryptosporidium parvum.
FH Key location/Qualifiers
FT CDS 1272..3296
FT /tag= a
FT /transl_except= (pos:1536..1538, aa:His)
FT /transl_except= (pos:1578..1580, aa:Leu)
FT /transl_except= (pos:1659..1661, aa:Lys)
FT /transl_except= (pos:2034..2036, aa:Lys)
FT /transl_except= (pos:2265..2267, aa:Leu)
FT /transl_except= (pos:2931..2933, aa:Lys)

FT /transl_except= (pos:3093..3095, aa:Lys)
FT /transl_except= (pos:3105..3107, aa:His)
FT /transl_except= (pos:3267..3269, aa:Lys)

PN WO9742349-A1.
PD 13-NOV-1997.
PF 08-MAY-1997; U07972.
PR 09-MAY-1996; US-647351.
PA (METR-) METROPOLITAN WATER DISTRICT SOUTHERN CALIFORNIA.
PI De Leon R, Rochelle PA;
DR WPI; 97-558999/51.
DR P-PSDB; W31345.
PT Amplifying and detecting viable and infective protozoan oocysts -
PT by monitoring presence or absence of heat shock protein genes in
PT their transcripts
PS Claim 12; Page 23-26; 44pp; English.
CC This genomic DNA sequence comprises the heat shock protein HSP70
CC gene of Cryptosporidium parvum oocysts. The invention provides
CC processes and kits for detecting encysted forms of protozoa,
CC particularly Cryptosporidium (especially Cryptosporidium muris,
CC Cryptosporidium baileyi, Cryptosporidium wairi and C. parvum)
CC and Giardia, that are viable and infectious by the enzymatic
CC amplification (e.g. by PCR) of a target gene sequence such as HSP70.
CC A preferred primer pair (see T89991-92) can amplify DNA or RNA
CC from many different Cryptosporidium species, and can be used with a
CC probe (see T89997) that confirms the identity of the amplification
CC product. A most preferred primer pair (see T89993-94) is specific
CC to C. parvum and can also be used with a probe (see T89998) to
CC confirm sequence identity. In addition, both Cryptosporidium and
CC Giardia can be detected simultaneously using a multiplex
CC amplification reaction. The methods exploit the speed, sensitivity
CC and specificity associated with an amplification procedure, and
CC enable pathogenic forms of protozoa present in low copy numbers to
CC be identified and distinguished from morphologically similar, but
CC non-pathogenic protozoa.
SQ Sequence 3607 BP; 1265 A; 514 C; 716 G; 1112 T;

Query Match 9.4%; Score 36; DB 1; Length 3607;
Best Local Similarity 48.3%; Pred. No. 5.8;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAAACAACAACAAAAACAATTCATTCAGAAAAATATCTTAGGACTGATAT 60
DB 221 AATAAAAAATAAATATACACTTTTATTTAATATAAATCTAATATAAAGTGAATA 162
OY 61 TGGTAATATGTCATTTAATWRTTKTGGGCATTCCTTACATTTGCTTGACAAGA 120
DB 161 TGATCTCCTCATCAATCAATAATATTTTACAGACTAATCAACAATATGATATATA 102
OY 121 TTAAGATGCTGTGCCAAATTTGTATTTTATTTGGAGACTCTTATCAAAAGTAATGC 180
DB 101 TGAACATAATGCTGTGAATAATTAATTTTCATAGCAAAAAAATATTTTATTAATATTAC 42

RESULT 14
V21209_05
Continuation (6 of 17) of V21209 from base 500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000

| | | | |
|----|-----------|---------|---------|
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

| | | | | |
|---------------------------|--------|-----------------|-----------|----------------|
| Query Match | 9.38; | Score 35.8; | DB 1; | Length 110000; |
| Best Local Similarity | 43.08; | Pred. No. 9.5; | | |
| Matches 154; Conservative | 3; | Mismatches 201; | Indels 0; | Gaps 0; |

[illegible]

RESULT 15

X12949/c
ID X12949 standard; DNA; 9047 BP.

DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:12.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
RT vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998

PF 04-MAY-1998; U08985

PR 14-NOV-1997; US-066009-

PR 06-MAY-1997; US-044031.

PR 16-MAY-1997; US-046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PL Barash SC, Dillon
DD WDT 00-045371/04

DR WPL; 99-0451/1/04.

P1 New isolated Enterococcus faecalis polynucleotides and polypeptides

use to develop products for the detection of *Enterococcus*

use in vaccination

```

..      *****
..      C:\tm 1: Page 305-310: 2084ms: End of file
PS

```

A computer readable medium has been developed which has recorded on it

982 nucleotide sequences isolated from the *Enterococcus faecalis* genome

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the *Enterococcus faecalis* genome with

CC commercial importance. The products can be used to detect the presence

CC of *Enterococcus faecalis* in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or

CC another related organism, in vivo or in vitro. In particular the

polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

SQ Sequence 9047 BP; 3107 A; 1542 C; 1841 G; 2551 T;

| | | | | |
|--------------------------|--------|----------------|-----------|--------------|
| Query Match | 9.3% | Score 35.6; | DB 1; | Length 9047; |
| Best Local Similarity | 50.6%; | Pred. No. 7.9; | | |
| Matches 80; Conservative | 2; | Mismatches 76; | Indels 0; | Gaps 0; |

QY 53 ACIGATATTGGTAATATATGCGCAATTTAATWRTTRTKTGGGCATTTCTTACATTGTCT 112
| | | | | | | | : | | | | | |
Db 6436 AGTCACCTATTAAATGCCGTTTCTATGAATAATAATCAAGGTGCCATTACTTTGATGGAT 63177

QY 113 TGACAAGATTAANAATGTCGTGCCAAAATTTTGATTTATTTGAGACTTCTATCAA 172
| | | | | | | | | | | | | | | | | |
Db 6376 TGAATATTGGAACACTTTCGACAAAATCTAAGAATTTATTGGCATGTTTTACAAA 63172

QY 173 AGTAATGCTGCCAAGAAGATCTAAGGAATTAGTAGTG 210
| | | | | | | | | | | | | | | | | |
Db 6316 CACTTGGCTATTGGAAGGAACCGTAGCAGATAATATTG 6279

Search completed: September 25, 1999, 09:58:34
Job time: 8165 sec

OY 175 TAATGCTGCCAAAGAGTCTAAGATAGTAGTCTCCMTCACCTGTTGGAGTGTG 234
|||
Db 1919 GAACACACAAAACTGACATATAAGGATATAGTGTTCAGGTTACAGGTACAGGGTTT 1860
OY 235 CTATCTTAAAGATTTGATTTCTCTGGAATGACATATATTTTACTTTGGTG 288
|||
Db 1859 CTTTTTCAAGGCTGTATGAATCTGCTCCTCCTTGAGATTTTAAGATTCTTG 1806

RESULT 2

US-08-647-351B-1/c
; Sequence 1, Application US/08647351B
; Patent No. 5770368

; GENERAL INFORMATION:

; APPLICANT: De Leon, Ricardo
; APPLICANT: Rochelle, Paul
; TITLE OF INVENTION: Cryptosporidium Detection Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheldon & Mak
; STREET: 225 S. Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: California
; ZIP: 91101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows version 6.1
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/647,351B
; FILING DATE: May 9, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: Farah, David A.
; REGISTRATION NUMBER: 38,134
; REFERENCE/DOCKET NUMBER: 11364
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3607 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double stranded

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

US-08-647-351B-1

Query Match 9.4%; Score 36; DB 3; Length 3607;
Best Local Similarity 48.3%; Pred. No. 0.72;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAACAACAATCTTCATTCAGAAAAATTATCTTAGGACTGATAT 60
|||
Db 221 AATAAAAAAATAATACAACTTTATTTATAAAAAATCTAATAAATAAAGTGAATA 162
OY 61 TGGTATATATGTCATTTATATWRTTKTGGGCATTTCTTACATGCTTGACAGA 120
|||
Db 161 TGGATCTCCTCATCAATCAATAATATTTTACAGACTAATCAACAATATGATTATTA 102
OY 121 TTAAGATGCTGTCACAAATTTGATTTTATTTGGAGACTTCTTATCAAAAGTAATGC 180
|||
Db 101 TGAACATAATGCTGTAAATTAATTTCAATAGGAAAAAATATTTTATATATATTAC 42

RESULT 3

US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:

; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:

; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5852 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 2378..5038

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 2378..5038

US-07-867-106-2

Query Match 9.1%; Score 34.8; DB 1; Length 5852;
Best Local Similarity 48.9%; Pred. No. 1.6;
Matches 87; Conservative 2; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAACAACAATCTTCATTCAGAAAAATTATCTTAGGACTGATAT 60
|||
Db 2022 AAAAAAAAAAAAAAAAAAACCCTCATTTATTAATATTAATTAATTTGTTTGA 2081
OY 61 TGGTATATATGTCATTTATATWRTTKTGGGCATTTCTTACATGCTTGACAGA 120
|||
Db 2082 TTTTATTTTATTAATAATTTAAATTTTATTTCTCTATCTAATTAATACCTTATTTAATAATA 2141
OY 121 TTAAGATGCTGTCACAAATTTGATTTTATTTGGAGACTTCTTATCAAAAGTAAT 178
|||
Db 2142 TTGATATATATCAAAATATTTATCAGTTTGGCATGACAAATTTAATTAATATTAATT 2199

RESULT 4

US-08-062-632-4/c
; Sequence 4, Application US/08062632
; Patent No. 5712090
; GENERAL INFORMATION:
; APPLICANT: Artushin, Sergey
; APPLICANT: Stipkovits, Laslo

RESULT 11
US-08-832-883-65/C
Sequence 65, Application US/08832883
Patent No. 5807681
GENERAL INFORMATION:
APPLICANT: Giordano, Antonio
APPLICANT: Baldi, Alphonso
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
TITLE OF INVENTION: OF CANCER
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,883
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 901 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-883-65

Query Match 8.1%; Score 31.2; DB 3; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2;
Matches 114; Conservative 4; Mismatches 122; Indels 2; Gaps 1;
QY 20 AACAACTTCTCATTCAGAAAAATATCTTAGGAGCTGATATTGTAATTATGTCATTT 79
DB 774 AATAACTTAACCTTAAGAAAAATATCTTAATAGCAAACTTAAGTCTTAGTTGTGCCA 715
QY 80 AATWRTTTRTKTGGGCGCATTTCCCTTACAT--TGCTTGACAGATTAAATGCTGTGCCA 137
DB 714 GTTATTATTGTGAAGCAATTTTATATACATTATGTCATGTCATCTTACACAACCCCATAT 655
QY 138 AAATTTTGTATTTTATTGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAGTCTAA 197
DB 654 GCTAGGAAGTATTATATTTCCCATTTATATATAGAGAAATGAGGTACAGAGAAATTTA 595
QY 198 GGAATTAGTAGTGTCCMTCACCTGTTTGAGAGTGTGCTATTCTAAAGATTGATTTTC 257
DB 594 ATGATTTGCAGGGTTACCAACCACTGTTAAGTACTGGGAAATTTGAACCCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 12
US-08-832-877-65/C
Sequence 65, Application US/08832877
Patent No. 5840506

GENERAL INFORMATION:
APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/832,877
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 901 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-877-65

Query Match 8.1%; Score 31.2; DB 4; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2;
Matches 114; Conservative 4; Mismatches 122; Indels 2; Gaps 1;
QY 20 AACAACTTCTCATTCAGAAAAATATCTTAGGAGCTGATATTGTAATTATGTCATTT 79
DB 774 AATAACTTAACCTTAAGAAAAATATCTTAATAGCAAACTTAAGTCTTAGTTGTGCCA 715
QY 80 AATWRTTTRTKTGGGCGCATTTCCCTTACAT--TGCTTGACAGATTAAATGCTGTGCCA 137
DB 714 GTTATTATTGTGAAGCAATTTTATATACATTATGTCATGTCATCTTACACAACCCCATAT 655
QY 138 AAATTTTGTATTTTATTGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAGTCTAA 197
DB 654 GCTAGGAAGTATTATATTTCCCATTTTATATATAGAGAAATGAGGTACAGAGAAATTTA 595
QY 198 GGAATTAGTAGTGTCCMTCACCTGTTTGAGAGTGTGCTATTCTAAAGATTGATTTTC 257
DB 594 ATGATTTGCAGGGTTACCAACCACTGTTAAGTACTGGGAAATTTGAACCCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 13
US-08-480-604A-9
Sequence 9, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.

: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
:
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,604A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/422,711
: FILING DATE: 14-APR-1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/405,496
: FILING DATE: 16-MAR-1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
:
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7101 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..7098
:
: US-08-480-604A-9

Query Match 8.0%; Score 30.8; DB 3; Length 7101;
Best Local Similarity 51.8%; Pred. No. 15;
Matches 59; Conservative 4; Mismatches 51; Indels 0; Gaps 0;

QY 8 ACAACAAAAAACAATCTTCATTCAGAAAAATATCTTAGGCACTGATATTGTAAT 67
DB 4227 AGAAGGAATAATGCAATTAAGAAGTTGATTTATTCATAATCAATAATTACTTAT 4286
QY 68 TATGTCATTAATWTRTRTKTGCGGCAATTCCTTACATGTCTTGACAAGAT 121
DB 4287 TTCTGGCAATTAATAATATTTGATGTTAAATTCAAATCATATTCACAGAAAAAT 4340

RESULT 14
US-08-377-687-48/C
; Sequence 48, Application US/08377687

: Patent No. 5538525
: GENERAL INFORMATION:
: APPLICANT: BROEKAERT, WILLEM F.
: APPLICANT: CAMUE, BRUNO P.A.
: APPLICANT: OSBORN, RUPERT W.
: APPLICANT: REES, SARAH B.
: APPLICANT: TERRAS, FRANKY R.G.
: APPLICANT: VANDERLEYDEN, JOZEF
: TITLE OF INVENTION: BIOCIDAL PROTEINS
: NUMBER OF SEQUENCES: 59
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DARBY & CUSHMAN
: STREET: 1100 NEW YORK AVENUE, N.W.
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/377,687
: FILING DATE:
: CLASSIFICATION: 800
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/002,480
: FILING DATE: 04-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: KORULIS, PAUL N.
: REGISTRATION NUMBER: 16,773
: REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-822-0944
: TELEFAX: 202-822-0944
: INFORMATION FOR SEQ ID NO: 48:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 414 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 16..255
:
: US-08-377-687-48

Query Match 8.0%; Score 30.6; DB 1; Length 414;
Best Local Similarity 46.9%; Pred. No. 9.8;
Matches 84; Conservative 3; Mismatches 92; Indels 0; Gaps 0;

QY 20 AACAAATCTTCATTCAGAAAAATTAATCTTAGGCACTGATATTGTAATTAATGTCATTT 79
DB 391 AATAAACTTTATTTGTATACCGAACCAACATACATATCTGTACATGCTTAAATC 332
QY 80 AATWTRTRTKTGCGGCAATTCCTTACATGTCTTGACAAGATTAATGTCGTGCCAA 139
DB 331 ACTCATGATAGTAGACACTGACTTATTTGTGTAATTAATCAATAAACTATTCACCAAG 272
QY 140 ATTTGTATTTTATTTGGAGACTTCTTATCAAAAGTATGCTGCCAAAGAGTCTAAG 198
DB 271 AGTTGGCATTAATTAACAAGAGAGTAGACAGATACACTTGTGAGCTGGAGACATAG 213

RESULT 15
US-08-206-176-3
; Sequence 3, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A


```

1  APPLICANT: Prunkard, Donna E
2  APPLICANT: Foster, Donald C
3  TITLE OF INVENTION: Production of Fibrinogen in Transgenic
4  TITLE OF INVENTION: Animals
5  NUMBER OF SEQUENCES: 27
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: ZymoGenetics, Inc.
8  STREET: 4225 Roosevelt Way, N.E.
9  CITY: Seattle
10 STATE: WA
11 COUNTRY: USA
12 ZIP: 98105
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/206,176
20 FILING DATE:
21 CLASSIFICATION: 435
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Parker, Gary E
24 REGISTRATION NUMBER: 31-648
25 REFERENCE/DOCKET NUMBER: 93-15
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 206-547-8080 ext 322
28 TELEFAX: 206-548-2329
29 INFORMATION FOR SEQ ID NO: 3:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 8878 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 IMMEDIATE SOURCE:
37 CLONE: human fibrinogen B-beta chain
38 FEATURE:
39 NAME/KEY: misc_RNA
40 LOCATION: 1..469
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 470..583
44 FEATURE:
45 NAME/KEY: intron
46 LOCATION: 584..3257
47 FEATURE:
48 NAME/KEY: exon
49 LOCATION: 3258..3449
50 FEATURE:
51 NAME/KEY: intron
52 LOCATION: 3450..3938
53 FEATURE:
54 NAME/KEY: exon
55 LOCATION: 3939..4122
56 FEATURE:
57 NAME/KEY: intron
58 LOCATION: 4123..5042
59 FEATURE:
60 NAME/KEY: exon
61 LOCATION: 5043..5270
62 FEATURE:
63 NAME/KEY: intron
64 LOCATION: 5271..5830
65 FEATURE:
66 NAME/KEY: exon
67 LOCATION: 5831..5944
68 FEATURE:
69 NAME/KEY: intron
70 LOCATION: 5945..6632
71 FEATURE:
72 NAME/KEY: exon
73 LOCATION: 6633..6758

```

```

? FEATURE:
? NAME/KEY: Intron
? LOCATION: 6759..6966
? FEATURE:
? NAME/KEY: exon
? LOCATION: 6967..7252
? FEATURE:
? NAME/KEY: Intron
? LOCATION: 7253..7870
? FEATURE:
? NAME/KEY: exon
? LOCATION: 7871..8102
? FEATURE:
? NAME/KEY: 3'UTR
? LOCATION: 8103..8537
? FEATURE:
? NAME/KEY: misc_RNA
? LOCATION: 8538..8878
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(470..583, 3258..3449, 3939..4122, 5043..5270
? : 5831..5944, 6633..6758, 6967..7252, 7871..8102)
US-08-206-176-3

```

| | | | | |
|-----------------------|--------|---------------|-------|----------------|
| Query Match | 8.0%; | Score 30.6; | DB 1; | Length 8878; |
| Best Local Similarity | 52.3%; | Pred. No. 18; | | |
| Matches | 57; | Conservative | 4; | Mismatches 48; |
| | | | | Indels 0; |
| | | | | Gaps 0; |

[illegible]

Search completed: September 25, 1999, 07:30:14
Job time: 3451 sec

```

1  APPLICANT: Prunkard, Donna E
2  APPLICANT: Foster, Donald C
3  TITLE OF INVENTION: Production of Fibrinogen in Transgenic
4  TITLE OF INVENTION: Animals
5  NUMBER OF SEQUENCES: 27
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: ZymoGenetics, Inc.
8  STREET: 4225 Roosevelt Way, N.E.
9  CITY: Seattle
10 STATE: WA
11 COUNTRY: USA
12 ZIP: 98105
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/206,176
20 FILING DATE:
21 CLASSIFICATION: 435
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Parker, Gary E
24 REGISTRATION NUMBER: 31-648
25 REFERENCE/DOCKET NUMBER: 93-15
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 206-547-8080 ext 322
28 TELEFAX: 206-548-2329
29 INFORMATION FOR SEQ ID NO: 3:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 8878 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 IMMEDIATE SOURCE:
37 CLONE: human fibrinogen B-beta chain
38 FEATURE:
39 NAME/KEY: misc_RNA
40 LOCATION: 1..469
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 470..583
44 FEATURE:
45 NAME/KEY: intron
46 LOCATION: 584..3257
47 FEATURE:
48 NAME/KEY: exon
49 LOCATION: 3258..3449
50 FEATURE:
51 NAME/KEY: intron
52 LOCATION: 3450..3938
53 FEATURE:
54 NAME/KEY: exon
55 LOCATION: 3939..4122
56 FEATURE:
57 NAME/KEY: intron
58 LOCATION: 4123..5042
59 FEATURE:
60 NAME/KEY: exon
61 LOCATION: 5043..5270
62 FEATURE:
63 NAME/KEY: intron
64 LOCATION: 5271..5830
65 FEATURE:
66 NAME/KEY: exon
67 LOCATION: 5831..5944
68 FEATURE:
69 NAME/KEY: intron
70 LOCATION: 5945..6632
71 FEATURE:
72 NAME/KEY: exon
73 LOCATION: 6633..6758

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:32 ; Search time 1811.29 Seconds
(without alignments)
417.095 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383

Sequence: 1 AAAACAACAACAACAAAAA.....ATTAGCTATATGTTAAA 383

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 376.8 | 98.4 | 400 | 36 | AA653016 | AA653016 ns68g01.s |
| 2 | 374 | 97.7 | 387 | 35 | AA579247 | AA579247 nf36d10.s |
| 3 | 369.8 | 96.6 | 522 | 26 | W56075 | W56075 zc56b05.r1 |
| 4 | 347.2 | 90.7 | 469 | 43 | AI188401 | AI188401 qd08e02.x |
| 5 | 339.4 | 88.6 | 432 | 35 | C75566 | C75566 C75566 Huma |
| 6 | 334.4 | 87.3 | 472 | 41 | AI027536 | AI027536 ow52c09.x |
| 7 | 327.4 | 85.5 | 389 | 39 | AA879150 | AA879150 nw84e05.s |
| 8 | 320.6 | 83.7 | 442 | 43 | AI160743 | AI160743 qb49d03.x |
| 9 | 317.4 | 82.9 | 444 | 47 | AI522136 | AI522136 tl78f03.x |
| 10 | 313.4 | 81.8 | 419 | 35 | C75584 | C75584 C75584 Huma |
| 11 | 303.4 | 79.2 | 496 | 41 | AI005274 | AI005274 ou07f12.x |
| 12 | 291.6 | 76.1 | 473 | 41 | AI051146 | AI051146 oy49d04.x |
| 13 | 288 | 75.2 | 407 | 36 | AA622892 | AA622892 np58c12.s |
| 14 | 280.4 | 73.2 | 344 | 28 | C14656 | C14656 C14656 Clon |
| 15 | 277 | 72.3 | 463 | 46 | AI420653 | AI420653 tf13b04.x |
| 16 | 261.2 | 68.2 | 324 | 31 | AA318914 | AA318914 EST21101 |
| 17 | 257 | 67.1 | 442 | 45 | AI382141 | AI382141 te30c04.x |
| 18 | 229.6 | 59.9 | 231 | 36 | AA653045 | AA653045 ns71f07.s |
| 19 | 227 | 59.3 | 268 | 28 | AA093999 | AA093999 cl1544.se |
| 20 | 213 | 55.6 | 225 | 36 | AA652561 | AA652561 ns73a06.s |
| 21 | 203 | 53.0 | 332 | 36 | AA657851 | AA657851 nu08d07.s |
| 22 | 189.6 | 49.5 | 352 | 28 | AA093469 | AA093469 kk5000.se |
| 23 | 185.4 | 48.4 | 246 | 36 | AA641128 | AA641128 nr28h07.r |
| 24 | 185 | 48.3 | 372 | 43 | AI214972 | AI214972 gm31f09.x |
| 25 | 181.4 | 47.4 | 200 | 32 | AA372885 | AA372885 EST84852 |
| 26 | 171 | 44.6 | 373 | 35 | AA580097 | AA580097 nh51h04.s |
| 27 | 158.4 | 41.4 | 282 | 30 | AA216139 | AA216139 hp0520.se |
| 28 | 151.4 | 39.5 | 255 | 24 | H83872 | H83872 yv96e01.s1 |
| 29 | 130.8 | 34.2 | 410 | 21 | T61899 | T61899 yb96e01.s1 |
| 30 | 118.2 | 30.9 | 703 | 48 | AI546846 | AI546846 PN2.1_07_ |
| 31 | 102.8 | 26.8 | 110 | 34 | AA503672 | AA503672 ng78h10.s |
| 32 | 101.2 | 26.4 | 309 | 21 | T61960 | T61960 yb96e01.r1 |
| 33 | 92.4 | 24.1 | 576 | 39 | C85936 | C85936 C85936 Mous |
| 34 | 92.4 | 24.1 | 438 | 44 | AI313916 | AI313916 uj38g06.x |
| 35 | 92.4 | 24.1 | 373 | 46 | AI450559 | AI450559 mq79f08.x |
| 36 | 91 | 23.8 | 425 | 29 | AA137503 | AA137503 mq79f07.r |
| 37 | 83.2 | 21.7 | 231 | 36 | AA658025 | AA658025 nu15d01.s |
| 38 | 76 | 19.8 | 275 | 49 | AV003262 | AV003262 AV003262 |
| 39 | 68.6 | 17.9 | 210 | 49 | AV010724 | AV010724 AV010724 |
| 40 | 64 | 16.7 | 298 | 50 | AV044015 | AV044015 AV044015 |
| 41 | 62.4 | 16.3 | 194 | 45 | AI350423 | AI350423 qt17d05.x |
| 42 | 42.6 | 11.1 | 427 | 42 | AI152967 | AI152967 ud54d06.r |
| 43 | 42.6 | 11.1 | 437 | 46 | AI428145 | AI428145 ml45c02.x |
| 44 | 41.2 | 10.8 | 412 | 44 | AI289102 | AI289102 qw21g01.x |
| 45 | 40.8 | 10.7 | 392 | 23 | D61957 | D61957 HUM226F10B |

ALIGNMENTS

RESULT 1
AA653016
LOCUS AA653016 400 bp mRNA
DEFINITION ns68g01.s1 NCI_CGAP_Pt2 Homo sapiens CDNA clone IMAGE:1188816, mRNA
ACCESSION AA653016
NID 92584668
VERSION AA653016.1 GI:2584668

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 400)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1325236.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 678 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 376.
Location/Qualifiers
1.400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:118816"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMPI0; site_1: NotI; site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMPI0 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

BASE COUNT 130 a 55 c 70 g 145 t
ORIGIN

Query Match 98.4%; Score 376.8; DB 36; Length 400;
Best Local Similarity 97.9%; Pred. No. 3.4e-68;
Matches 375; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAACAACAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGACTGATAT 60
|||||
Db 14 AAACAACAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGACTGATAT 73
|||||

QY 61 TCGTAATTATGTCATTTAATWRTTKTGGGCACTTCCTTACATTCCTTGACAAGA 120
|||||
Db 74 TCGTAATTATGTCATTTAATATATTTGGGCACTTCCTTACATTCCTTGACAAGA 133
|||||

QY 121 TTAATAATGTCTGCCAATAATTTGTATTTATTTGAGACTCTTATCAAAAGTAATGC 180
|||||
Db 134 TTAATAATGTCTGCCAATAATTTGTATTTATTTGAGACTCTTATCAAAAGTAATGC 193
|||||

QY 181 TGCCAAAGGAAGCTAAGCAATTAGTAGTGTCCMTCACCTGTTGGAGTGTCTATTC 240
|||||
Db 194 TGCCAAAGGAAGCTAAGCAATTAGTAGTGTCCMTCACCTGTTGGAGTGTCTATTC 253
|||||

QY 241 TMAAAGATTTTGATTTCTGGAATGACAAATTATATTTAACCTTTGGTGGGGAANAAGTT 300
|||||

|||||
Db 254 TAAAGATTTTGATTTCTCGAATGACAAATTATATTTTAACCTTTGGTGGGGAAGAGTT 313
|||||

QY 301 ATAGGACCACAGCTCTCAGCTTCTGATACCTGTAATTAATCTTTATTTGCACCTGTTTG 360
|||||

Db 314 ATAGGACCACAGCTCTCAGCTTCTGATACCTGTAATTAATCTTTATTTGCACCTGTTTG 373
|||||

QY 361 ACCATTAAGCTATATAGTTTAAA 383
|||||

Db 374 ACCATTAAGCTATATAGTTTAAA 396
|||||

RESULT 2
AA579247 387 bp mRNA EST 12-SEP-1997
LOCUS n36d10.s1 NCI_CGAP_Pr2 Homo sapiens CDNA clone IMAGE:915859, mRNA
DEFINITION sequence.
ACCESSION AA579247
NID 92357431
VERSION AA579247.1 GI:2357431
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 387)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 681 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 351.
Location/Qualifiers
1.387
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915859"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMPI0; site_1: NotI; site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMPI0 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

BASE COUNT 125 a 54 c 67 g 141 t
ORIGIN

Query Match 97.7%; Score 374; DB 35; Length 387;
Best Local Similarity 98.4%; Pred. No. 1.3e-67;

Matches 371; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAACAACAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATAT 60
|||||
Db 11 AAACAACAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATAT 70
OY 61 TGGTAATATGTCATTTAATWTRTRTKGGGCATTTCTTACATGTCTTGACAAGA 120
|||||
Db 71 TGGTAATATGTCATTTAATAATATTTGGGCATTTCTTACATGTCTTGACAAGA 130
OY 121 TTAAGATGTCGCAAAATTTGTATTTTATTTGGAGACTCTTATCAAAAGTATGC 180
|||||
Db 131 TTAAGATGTCGCAAAATTTGTATTTTATTTGGAGACTCTTATCAAAAGTATGC 190
OY 181 TGCCAAAGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTGTGAGTGTCTATTC 240
|||||
Db 191 TGCCAAAGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTGTGAGTGTCTATTC 250
OY 241 TAAAGATTTTGATTTCTCGAATGACAAATATATTTTAACTTTGGGGGAAANAGTT 300
|||||
Db 251 TAAAGATTTTGATTTCTCGAATGACAAATATATTTTAACTTTGGGGGAAAGAGTT 310
OY 301 ATAGACACACAGTCTTCACCTCTGATACCTGTAAATTAATCTTTATTCACCTGTTTG 360
|||||
Db 311 ATAGACACACAGTCTTCACCTCTGATACCTGTAAATTAATCTTTATTCACCTGTTTG 370
OY 361 ACCATTAAGCTATATGT 377
|||||
Db 371 ACCATTAAGCTATATGT 387

RESULT 3
LOCUS W56075 522 bp mRNA EST 11-OCT-1996
DEFINITION zc56b05.r1 Soares_parathyroid_tumor_NbHRA Homo sapiens cDNA clone
IMAGE:326289 5', mRNA sequence.
ACCESSION W56075
NID 91357965
VERSION W56075.1 GI:1357965
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Mairra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785211.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 595 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 335.
Location/Qualifiers
1..522
/organism="Homo sapiens"
/db_xref="GDB:1260491"
/db_xref="taxon:9606"
/map="643H05"
/clone="IMAGE:326289"
/clone_lib="Soares_parathyroid_tumor_NbHRA"

/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer
[5',
TGTACCAATCTGAGTGGAGCGCGCCGACCAATTTTCTTTTCTTTTCTTTT
T-3']], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

Query Match 96.6%; Score 369.8; DB 26; Length 522;
Best Local Similarity 97.9%; Pred. No. 8.6e-67;
Matches 368; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 7 AACAAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATATGCTAA 66
|||||
Db 1 AACAAACAAAAAACAATCTTCATTCAGAAAAATTATCTTAGGGACTGATATGCTAA 60
OY 67 TTATGTCATTTAATWTRTRTKGGGCATTTCTTACATGTCTTGACAAGATTAATA 126
|||||
Db 61 TTATGTCATTTAATATATATTTTGGGCATTTCTTACATGTCTTGACAAGATTAATA 120
OY 127 TGTCTGTGCCAAATTTTGTATTTTATTTGGAGACTCTTATCAAAAGTATGTCGCAA 186
|||||
Db 121 TGTCTGTGCCAAATTTTGTATTTTATTTGGAGACTCTTATCAAAAGTATGTCGCAA 180
OY 187 AGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTGTGAGTGTGCTATCTAAAG 246
|||||
Db 181 AGGAAGTCTAAGGAATAGTAGTGTCCCMTCACCTGTGTGAGTGTGCTATCTAAAG 240
OY 247 ATTTGATTTCTCGAATGACAAATTAATTTTAACTTTGGGGGGAANAGTTATAGGA 306
|||||
Db 241 ATTTGATTTCTCGAATGACAAATTAATTTTAACTTTGGGGGGAANAGTTATAGGA 300
OY 307 CCACAGTCTTCACCTCTGATACCTGTAAATTAATCTTTATTCACCTGTTTGACATT 366
|||||
Db 301 CCACAGTCTTCACCTCTGATACCTGTAAATTAATCTTTATTCACCTGTTTGACATT 360
OY 367 AAGCTATATGTTTAA 382
|||||
Db 361 AAGCTATATGTTTAA 376

RESULT 4
LOCUS A1188401/c 469 bp mRNA EST 28-OCT-1998
DEFINITION qd08e02.x1 Soares_placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
clone IMAGE:1723130 3', mRNA sequence.
ACCESSION A1188401
NID 93739610
VERSION A1188401.1 GI:3739610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 469)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1899479.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 564 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 444.

FEATURES
source
1. .472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1650448"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer
[5'-
TGTACCATCTGAAGTGGAGCGCGCCGACCAATTTTCTTTTCTTTTCTTTT
T-3']
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 175 a 80 c 55 g 162 t
ORIGIN

Query Match 87.3%; Score 334.4; DB 41; Length 472;
Best Local Similarity 97.9%; Pred. No. 1.5e-59;
Matches 332; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 45 TCTTAGGAGCTGATATGCTATATGTCATTTATATWRTKTTGGGCATTCCTTA 104
|||||
DB 472 TCTTAGGAGCTGATATGCTATATGTCATTTATATATATTTTGGGCATTCCTTA 413
OY 105 CATGTCTTGACAGATTAAATGCTGCGCCAAAATTTGTAATTTATTGGAGACTTC 164
|||||
DB 412 CATGTCTTGACAGATTAAATGCTGCGCCAAAATTTGTAATTTATTGGAGACTTC 353
OY 165 TTATCAAAAGTAAATGCTGCCAAAGAGCTAAAGAAATTAGTAGTCTCCMTCACCTGT 224
|||||
DB 352 TTATCAAAAGTAAATGCTGCCAAAGAGCTAAAGAAATTAGTAGTCTCCMTCACCTGT 293
OY 225 TTGAGTGTGCTATTTCTAAAGATTTTGAATTTCTGGAATGACAAATTAATTTAATCTT 284
|||||
DB 292 TTGAGTGTGCTATTTCTAAAGATTTTGAATTTCTGGAATGACAAATTAATTTAATCTT 233
OY 285 GGTGGGGAANAAGTTATAGGACACAGCTTCACTTCTGATACCTGTAATTAATCTTT 344
|||||
DB 232 GGTGGGGAANAAGTTATAGGACACAGCTTCACTTCTGATACCTGTAATTAATCTTT 173
OY 345 TATGCACTTGTTTGACCAATTAAAGCTAATGTTTAAA 383
|||||

DB 172 TATGCACTTGTTTGACCAATTAAAGCTATATGTTAGAA 134

RESULT 7
AA879150 389 bp mRNA EST 25-MAR-1998
LOCUS nw84e05.s1 NCI_CGAP_Pr12 Homo sapiens CDNA clone IMAGE:1253312,
DEFINITION mRNA sequence.
ACCESSION AA879150
NID 92988115
VERSION AA879150.1 GI:2988115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 389)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153434.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 381.
Location/Qualifiers
1. .389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone="IMAGE:1253312"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

FEATURES
source

BASE COUNT 121 a 51 c 75 g 142 t
ORIGIN

Query Match 85.5%; Score 327.4; DB 39; Length 389;
Best Local Similarity 95.5%; Pred. No. 4.2e-58;
Matches 342; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

OY 26 TCTTCATTCAGAAAAATTATCTTAGGGAGCTGATATGTAATTATGTCATTTAATWRT 85
|||
DB 3 TCGACATTCAGAAAAATTATCTTAGGGAGCTGATATGTAATTATGTCATTTAATWRT 62
OY 86 RTTKTGGGCAATTCCTTACATGCTCTGACAGATTAATGCTGTGCCAAATTTTG 145
:||
DB 63 ATT-TGGGCAATTCCTTACATGCTCTGACAGATTAATGCTGTGCCAAATTTTG 121
OY 146 TATTTAATTTGAGAGCTCTTATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAG 205
|||||
DB 122 TATTTAATTTGAGAGCTCTTATCAAAAGTAATGCTGCCAAAGAGCTAAGGAATTAG 181
OY 206 TAGTGTCCMTCACCTGTTTGGAGTGTGCTATTTCTAAAGATTTTGATTTCTCGAATG 265
|||||

| Db | 182 | 241 | 241 |
|------------|--|--|-----|
| QY | 266 | ACAAATTATATTTTAACTTTGGTGGGGGAAANAGTTATAGACCAAGCTTCACTTCTGA | 325 |
| Db | 242 | ACAAATTATATTTTAACTTTGGTGGGGGAAAGAGTTATAGACCAAGCTTCACTTCTGA | 301 |
| QY | 326 | TACTTGTAATTAATCTTTTATTCGACTTGTGTTGACCATTAAAGCTATATGTTTAAA | 383 |
| Db | 302 | TACTTGTAATTAATCTTTTATTCGACTTGTGTTGACCATTAAAGCTATATGTTTAAA | 359 |
| RESULT | 8 | | |
| LOCUS | AI160743/c | | |
| DEFINITION | AI160743 442 bp mRNA EST 26-OCT-1998 | | |
| ACCESSION | qB49d03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1703429 | | |
| NID | AI160743 | | |
| VERSION | g3694123 | | |
| KEYWORDS | AI160743.1 GI:3694123 | | |
| SOURCE | EST. | | |
| ORGANISM | human. | | |
| REFERENCE | Homo sapiens | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | |
| TITLE | Eutheria; Primates; Catarrhini; Homidae; Homo. | | |
| JOURNAL | 1 (bases 1 to 442) | | |
| COMMENT | NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . | | |
| | National Cancer Institute / National Institute of Neurological | | |
| | Disorders and Stroke, Brain Tumor Genome Anatomy Project | | |
| | (CGAP/BTGP), Tumor Gene Index | | |
| | Unpublished (1998) | | |
| | On Jan 19, 1998 this sequence version replaced gi:2286564. | | |

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.C.E. Consortium/LNL at:
www.bio.lnl.gov/bbrp/image/image.html

Insert Length: 2616 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amer sham
High quality sequence stop: 433.

FEATURES

source

```

source
1. .442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1703429"
/clone_1lb="NCI_CGAP_Brn23"
/tissue_type="g1ioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGGTGGAGCGGCCGACATCTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```

| | | | | |
|----------------------------|--------|------------------|--------|-------------|
| Query Match | 83.7%; | Score 320.6; | DB 43; | Length 442; |
| Best Local Similarity | 97.0%; | Pred. No. 1e-56; | | |
| Matches, 320; Conservative | 5; | Mismatches | 5; | Indels 0; |
| | | | | Gaps 0; |

| | | | |
|----|-----|---|-----|
| QY | 54 | CTGATATTGGTAATTATGGTCAATTTAAFWBRTKTTGGGGCAATTCCTTACATTGCTT | 113 |
| Db | 442 | CTGATATTGGTAATTATGGTCAATTTAATAATATTTTGAGGCAATTCCTTACATTGCTT | 383 |
| QY | 114 | GACAAGATTAAAAATGCTCTGTGCCAAAATTTTGTATTATTATTTGGAGACTTCTTATCAAAA | 173 |
| Db | 382 | GACAAGATTAAAAATGCTCTGTGCCAACATTTTGTATTATTATTTGGAGACTTCTTATCAAAA | 323 |
| QY | 174 | GTAATGCTGCCAAAGGAAGCTTAAGGAATTAGTAGTGTTCCTCCMTCACTGTTTGAGAGTGT | 233 |
| Db | 322 | GTAATGCTGCCAAAGGAAGCTTAAGGAATTAGTAGTGTTCCTCCATCACTGTTTGAGAGTGT | 263 |
| QY | 234 | GCTATTCTAAAGAATTTTGATTTCCGTGAATGACAATTATATTTAACCTTTGGTGGGGA | 293 |
| Db | 262 | GCTATTCTAAAGAATTTTGATTTCCGTGAATGACAATTATATTTAACCTTTGGTGGGGA | 203 |
| QY | 294 | AANAGTTATAGACACAGTCTTCACTTCTGATACTTGTAAATTAATCTTTTATTGCACT | 353 |
| Db | 202 | AAGAGTTATAGACACAGTCTTCACTTCTGATACTTGTAAATTAATCTTTTATTGCACT | 143 |
| QY | 354 | TGTTTTGACCATTAACTATATGTTTAAA | 383 |
| Db | 142 | TGTTTTGGCCATTAACTATATGTTTAAA | 113 |

| | | | | | | | | | |
|------------|----------------|----------------|--------------|------------|---------------|----|--|--|--|
| RESULT | 9 | | | | | | | | |
| LOCUS | AI522136/c | | | | | | | | |
| DEFINITION | AI522136 | 444 bp | mRNA | EST | 13-APR-1999 | | | | |
| | t178f03.x1 | NCI_CGAP_K1d11 | Homo sapiens | CDNA clone | IMAGE:2138141 | 3' | | | |
| | mRNA sequence. | | | | | | | | |
| ACCESSION | AI522136 | | | | | | | | |
| NID | 94436271 | | | | | | | | |
| VERSION | AI522136.1 | GI:4436271 | | | | | | | |
| KEYWORDS | EST. | | | | | | | | |
| SOURCE | human. | | | | | | | | |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 444)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
On Mar 10, 1998 this sequence version replaced gi:2948750.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found at <http://www.scripps.edu/~ncicga>

found through the I.M.A.G.E. Consortium

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 530 Std Error: 0.00

Insert Length: 539 Std Error: 0.00
Sed primer: -400bp from Ctpoc

seq primer: -400P from G1bco
Location/Qualifiers

FEATURES

source

4
6
1
4
6

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2138141"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAA
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs

```


from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo.

BASE COUNT 170 a 75 c 51 g 148 t
ORIGIN

Query Match 82.9%; Score 317.4; DB 47; Length 444;
Best Local Similarity 97.8%; Pred. No. 4.5e-56;
Matches 315; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 62 GGTAATTATGGCAATTTATWRTTRTGGGGCATTCCCTACATGTCCTGACAGAT 121
|||||
DB 444 GGTAATTATGGCAATTTATWRTTRTGGGGCATTCCCTACATGTCCTGACAGAT 385
|||||
OY 122 TAAATGTCGTGCCAAATTTGTATTTATTTGGAGACTTCTTCAAAAGTAATGCT 181
|||||
DB 384 TAAATGTCGTGCCAAATTTGTATTTATTTGGAGACTTCTTCAAAAGTAATGCT 325
|||||
OY 182 GCCAAGAAGTCTAAGCAATTAGTAGTCCCMTCACCTGTTGGAGTGTCTATCT 241
|||||
DB 324 GCCAAGAAGTCTAAGCAATTAGTAGTCCCMTCACCTGTTGGAGTGTCTATCT 265
|||||
OY 242 AAAGATTTGATTTCTCGATGACATTAATTTAATCTTTGCTGGGGAAGTTA 301
|||||
DB 264 AAAGATTTGATTTCTCGATGACATTAATTTAATCTTTGCTGGGGAAGTTA 205
|||||
OY 302 TAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATTTGACCTGTTTGA 361
|||||
DB 204 TAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATTTGACCTGTTTGA 145
|||||
OY 362 CCATTAGCTATATGTTTAAA 383
|||||
DB 144 CCATTAGCTATATGTTTAAA 123
|||||

RESULT 10
LOCUS C75584 419 bp mRNA EST 09-SEP-1997
C75584/c
DEFINITION C75584 Human pancreatic islet Homo sapiens cDNA clone hbc7897, mRNA
sequence.
ACCESSION C75584
NID 92366646
VERSION C75584.1 GI:2366646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 419)
AUTHORS Takeda, J.
TITLE Large scale collection of expressed sequence tags (ESTs) from human
pancreatic islet cDNA library
JOURNAL Unpublished (1997)
COMMENT On May 5, 1995 this sequence version replaced gi:797622.

CONTACT: Jun Takeda
Institute for Molecular and Cellular Regulation, Gunma University
3-39-15 Showa-machi, Maebashi Gunma 371, Japan
Tel: 272-20-8856
Fax: 272-20-8896
Email: jtakeda@sb.gunma-u.ac.jp.
LOCATION/Qualifiers
1. .419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hbc7897"
/clone_lib="Human pancreatic islet"
/note="Vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
I; mRNA was prepared from normal adult human islets. cDNA
was directionally synthesized from the Xho I in the vector
to the EcoRI site. cDNA was size fractionated to remove

sequences <1000 bp in size."
BASE COUNT 159 a 76 c 50 g 132 t 2 others
ORIGIN

Query Match 81.8%; Score 313.4; DB 35; Length 419;
Best Local Similarity 97.0%; Pred. No. 3e-55;
Matches 322; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 53 ACTGATATTGGTATATATGTCATTTATWRTTRTGGGGCATT-CCTTACATGTC 111
|||||
DB 419 ACTGATATTGGTATATATGTCATTTATWRTTRTGGGGCATTCCCTTACATGTC 360
|||||
OY 112 TTGACAGATTTAAATGTCGTGCCAAATTTGTATTTATTTGGAGACTTCTATCAA 171
|||||
DB 359 TTGACAGATTTAAATGTCGTGCCAAATTTGTATTTATTTGGAGACTTCTATCAA 300
|||||
OY 172 AAGTAATGCTGCCAAGAAGTCTAAGCAATTAGTAGTCCCMTCACCTGTTGAGT 231
|||||
DB 299 AAGTAATGCTGCCAAGAAGTCTAAGCAATTAGTAGTCCCMTCACCTGTTGAGT 240
|||||
OY 232 GTGCTATTCTAAAGATTTTGAATTTCTCGAATGACATTAATTTAATCTTTGAGG 291
|||||
DB 239 GTGCTATTCTAAAGATTTTGAATTTCTCGAATGACATTAATTTAATCTTTGAGG 180
|||||
OY 292 GAAANAGTTATAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATGCA 351
|||||
DB 179 GAAAGAGTTATAGGACCACAGTCTTCACTTCTGATACCTGTAATTAATCTTTATGCA 120
|||||
OY 352 CTGTTTGGACCATTAAGCTATATGTTTAAA 383
|||||
DB 119 CTGTTTGGACCATTAAGCTATATGTTTAAA 88
|||||

RESULT 11
LOCUS AI005274 496 bp mRNA EST 27-AUG-1998
AI005274/c
DEFINITION Ou07f12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
AI005274
ACCESSION AI005274
NID 93214784
VERSION AI005274.1 GI:3214784
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1636 Std Error: 0.00
Seq primer: primer name ambiguous
High quality sequence stop: 454.
LOCATION/Qualifiers
1. .496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1625615"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 193 a 79 c 57 g 167 t
ORIGIN

Query Match 79.2%; Score 303.4; DB 41; Length 496;
Best Local Similarity 97.5%; Pred. No. 3.2e-53;
Matches 312; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

QY 64 TAATATGTCATTTAATWRTTKTGGGCATTTCTTACATGCTTGACAAAGTTA 123
|||||
Db 496 TAATATGTCATTTAATATATTTGGGGCATTTCTTACATGCTTGACAA-ATTA 438
QY 124 AAATGCTGTGCCAAAATTTGTATTATTGGAGACTTCTTATCAAAAGTAATGCTGC 183
|||||
Db 437 AAATGCTGTGCCAAAATTTGTATTATTGGAGACTTCTTATCAAAAGTAATGCTGC 378
QY 184 CAAGGAAGTCTAAGGAATAGTAGTGTTCCTCCATCTGTTGGAGTGTCTATTCTAA 243
|||||
Db 377 CAAGGAAGTCTAAGGAATAGTAGTGTTCCTCCATCTGTTGGAGTGTCTATTCTAA 318
QY 244 AAGATTTTGTATTCCTGGAATGACAATTATATTAACTTTGGTGGGGAANAAGTTATA 303
|||||
Db 317 AAGATTTTGTATTCCTGGAATGACAATTATATTAACTTTGGTGGGGAANAAGTTATA 258
QY 304 GGACCACAGCTCTTCACCTTCGATACTTGTAAATTAATCTTTATTCACCTGTTTGACC 363
|||||
Db 257 GGACCACAGCTCTTCACCTTCGATACTTGTAAATTAATCTTTATTCACCTGTTTGACC 198
QY 364 ATTAAGCTATATGTTAAAA 383
|||||
Db 197 ATTAAGCTATATGTTAGAA 178

RESULT 12
AI051146 473 bp mRNA EST 24-SEP-1998
LOCUS
DEFINITION OY49d04.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1669159 3',
mRNA sequence.

ACCESSION AI051146
NID 93306680
VERSION AI051146.1 GI:3306680
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 473)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)

JOURNAL
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044831.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 2319 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 438.
Location/Qualifiers
1. 473

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1669159"
/clone_lib="NCI_CGAP_Brn23"
/issue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 181 a 77 c 56 g 159 t
ORIGIN

Query Match 76.1%; Score 291.6; DB 41; Length 473;
Best Local Similarity 98.6%; Pred. No. 8.4e-51;
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 89 KTGCGGCATTTCCCTACATGTGCTTGACAGATTAAATGTCGTGCCAAAATTTGTAT 148
:|||||
Db 473 TTGGGGCATTTCCCTACATGTGCTTGACAGATTAAATGTCGTGCCAAAATTTGTAT 414
QY 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGGAAGCTAAGGAATTAGTAG 208
|||||
Db 413 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAAGGAAGCTAAGGAATTAGTAG 354
QY 209 TGTTCCTCAGCTGTTTGGAGTGTGCTATCTTAAAGATTTGATTTCTGGAATGACA 268
|||||
Db 353 TGTTCCTCAGCTGTTTGGAGTGTGCTATCTTAAAGATTTGATTTCTGGAATGACA 294
QY 269 ATTATATTTTAACCTTGGTGGGGAANAAGTTATAGGACCACAGCTCTTCACCTCTGATAC 328
|||||
Db 293 ATTATATTTTAACCTTGGTGGGGAANAAGTTATAGGACCACAGCTCTTCACCTCTGATAC 234
QY 329 TTGTAATTAATCTTTTATTCGACTGTTTGGACCATTAAAGCTATATGTTAAAA 383
|||||
Db 233 TTGTAATTAATCTTTTATTCGACTGTTTGGACCATTAAAGCTATATGTTAGAA 179

RESULT 13
AA622892 407 bp mRNA EST 21-OCT-1997
LOCUS
DEFINITION np58c12.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1130518 3',
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION AA622892
NID 92526768
VERSION AA622892.1 GI:2526768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 407)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT On May 18, 1995 this sequence version replaced gi:810969.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbp/limage/limage.html

Insert length: 948 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 385.

FEATURES

SOURCE

1. 407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130518"
/clone_lib="NCI-CGAP_Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. This library is the normalized version of NCI-CGAP Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 157 a 72 c 50 g 128 t
ORIGIN

Query Match 75.2%; Score 288; DB 36; Length 407;
Best Local Similarity 99.0%; Pred. No. 4.7e-50;
Matches 288; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 93 GGCATTTCCTTACATGTCCTGACACAGATTAAATGCTGTGCCAAAATTTGTAATTTA 152
|||||
Db 407 GGCATTTCCTTACATGTCCTGACACAGATTAAATGCTGTGCCAAAATTTGTAATTTA 348
QY 153 TTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAAAGTCTAAGGAATAGTAGTGT 212
|||||
Db 347 TTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAAAGTCTAAGGAATAGTAGTGT 288
QY 213 CCCATCACTGTTGGAGTGTCTATTCTTAAAGATTGATTTCCCTGAATGACAATTA 272
|||||
Db 287 CCCATCACTGTTGGAGTGTCTATTCTTAAAGATTGATTTCCCTGAATGACAATTA 228
QY 273 TATTTTAACCTTGGTGGGGAANAGTATATAGACACAGCTCTTCACCTTCGATACTGT 332
|||||
Db 227 TATTTTAACCTTGGTGGGGAANAGTATATAGACACAGCTCTTCACCTTCGATACTGT 168
QY 333 AATTAATCTTTATTCGACTGTTTGACCATTAAGCTATATGTTAAA 383
|||||
Db 167 AATTAATCTTTATTCGACTGTTTGACCATTAAGCTATATGTTAAA 117

RESULT 14
C14656 344 bp mRNA EST 30-SEP-1996
LOCUS C14656 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens CDNA
DEFINITION clone GEN-078E02 5', mRNA sequence.
ACCESSION C14656
NID 91569363
VERSION C14656.1 GI:1569363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 344)

AUTHORS Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
Fujiwara et al. (1995)
JOURNAL Unpublished (1995)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407498.

Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES

SOURCE

1. 344
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-078E02"
/clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 138 a 44 c 50 g 108 t 4 others
ORIGIN

Query Match 73.2%; Score 280.4; DB 28; Length 344;
Best Local Similarity 96.5%; Pred. No. 1.7e-48;
Matches 276; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTAATCTTAGGACTGATAT 60
|||||
Db 59 AAACAACAACAACAAAAACAATCTTCATTCAGAAAAATTAATCTTAGGACTGATAT 118
QY 61 TCGTAATATGTCATATTTAATWRTTRTKGGGCATTTCCCTTACATGCTCTGACACAGA 120
|||||
Db 119 TCGTAATATGTCATATTTAATATATTTGGGGCATTTCCCTTACATGCTCTGACACAGA 178
QY 121 TTAATATGCTGTGCCAAAATTTGTATTTATTTGGAGACTTCTTATCAAAAGTAATGC 180
|||||
Db 179 TTAATATGCTGTGCCAAAATTTGTATTTATTTGGAGACTTCTTATCAAAAGTAATGC 238
QY 181 TGCCAAAGAAAGCTTAAGGAATAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC 240
|||||
Db 239 TGCCAAAGAAAGCTTAAGGAATAGTAGTGTCCCMTCACCTGTTGGAGTGTCTATTC 298
QY 241 TAAAGATTTGATTTCTCGAATGACAATATATTTTAACTTTGG 286
|||||
Db 299 TAAAGATTTGATTTCTCGAATGACAATATATTTCAACTTTGG 344

RESULT 15
A1420653/c 463 bp mRNA EST 28-MAR-1999
LOCUS A1420653 ttf13b04.x1 NCI-CGAP_Brn23 Homo sapiens CDNA clone IMAGE:2096047 3',
DEFINITION mRNA sequence.
ACCESSION A1420653
NID 94266584
VERSION A1420653.1 GI:4266584
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 463)
AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980399.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldó, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1420 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 449.
Location/Qualifiers

FEATURES

source
1. .463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2096047"
/clone_1lb="NCI CGAP_Brn23"
/tissue_type="gliblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pT7/3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGGAGCGGCCCAATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7/3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldó."

BASE COUNT 179 a 76 c 55 g 153 t
ORIGIN

Query Match 72.3%; Score 277; DB 46; Length 463;
Best Local Similarity 98.6%; Pred. No. 8.1e-48;
Matches 288; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

```
QY 92 GGCATTTCCTACATGCTGTGACAGATTAAATGCTGTGCCAAATTTGTATTTT 151
    |||
Db 463 GGCATTTCCTACATGCTGTGACAGATTAAATGCTGTGCCAAATTTGTATTTT 404

QY 152 ATTTGAGACTTCTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAGTGT 211
    |||
Db 403 ATTTGAGACTT-TTATCAAAAGTAATGCTGCCAAAGAGCTTAAGGAATTAGTAGTGT 345

QY 212 TCCCMTCACCTGTGTGAGTGTGCTATCTAAAGATTGATTTCTTGAATGACAATT 271
    |||
Db 344 TCCCATCAGCTGTGTGAGTGTGCTATCTAAAGATTGATTTCTTGAATGACAATT 285

QY 272 ATATTTTAAGTTGGTGGGGAANAGTTATAGGACCACAGTCTTCACTTCTGATACCTTG 331
    |||
Db 284 ATATTTTAAGTTGGTGGGGAAGAGTATAGGACCACAGTCTTCACTTCTGATACCTTG 225

QY 332 TAAATTAATCTTTTATGCACTTGTGACCATTAAAGCTAATATGTTTAAA 383
    |||
Db 224 TAAATTAATCTTTTATGCACTTGTGACCATTAAAGCTAATATGTTTAAA 173
```

Search completed: September 25, 1999, 12:05:37
Job time: 8141 sec

OM of: US-09-030-606-224 to: A_Geneseq_36:* out_format : pfs
Date: Sep 25, 1999 11:36 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09030606/runat_24091999.171616_29804/app-query.fasta.1
-DB=A_Geneseq_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-224
Query length: 320
Database: A_Geneseq_36:*
Database sequences: 18963
Database length: 23686106
Search time (sec): 185.540000

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|---------------------|------|-------|--------|--------|------|-------------------------------------|
| A_Geneseq_36:W75057 | - | 71.00 | 149.45 | 0.1621 | 467 | Human secreted protein encoded |
| A_Geneseq_36:Y13378 | - | 70.00 | 146.95 | 0.2235 | 467 | Amino acid sequence of protein |
| A_Geneseq_36:R88732 | + | 66.00 | 133.48 | 0.8853 | 663 | S.aureus topoisomerase IV grlB |
| A_Geneseq_36:Y12066 | - | 63.50 | 145.80 | 1.21 | 100 | Human 5' EST secreted protein S |
| A_Geneseq_36:R77168 | - | 63.50 | 138.89 | 1.45 | 202 | Arabidopsis condensin enzyme c |
| A_Geneseq_36:Y07052 | + | 62.50 | 132.12 | 2.24 | 312 | Renal cancer associated antigen |
| A_Geneseq_36:W97757 | + | 62.50 | 125.94 | 2.64 | 585 | S-region transfer activity-70 C |
| A_Geneseq_36:R20030 | - | 59.00 | 116.40 | 8.30 | 633 | Bacillus thuringiensis CryIIIB c |
| A_Geneseq_36:W88448 | - | 59.50 | 110.61 | 8.51 | 1296 | Caenorhabditis elegans NPC1 pr |
| A_Geneseq_36:W39159 | + | 57.50 | 124.16 | 9.90 | 196 | Clone PRBS3FH2910 #4.1 CFH rela |
| A_Geneseq_36:Y07068 | + | 58.00 | 118.11 | 10.23 | 412 | Renal cancer associated antigen |
| A_Geneseq_36:W39158 | + | 57.50 | 122.63 | 10.31 | 229 | Clone PRBS3FH2910 #3.1 CFH rela |
| A_Geneseq_36:W28143 | + | 57.00 | 126.21 | 10.66 | 140 | Amino acid sequence of ATP depe |
| A_Geneseq_36:W39156 | + | 57.50 | 120.31 | 10.97 | 290 | Human partial complement factor |
| A_Geneseq_36:R56285 | - | 57.50 | 115.75 | 12.37 | 461 | Sequence of human alpha-L-fucos |
| A_Geneseq_36:W14995 | + | 57.50 | 112.84 | 13.36 | 620 | Human c-fos induced growth fact |
| A_Geneseq_36:R04752 | - | 56.50 | 119.66 | 14.40 | 240 | Amino acid sequence of maize ze |
| A_Geneseq_36:R66452 | - | 57.00 | 112.45 | 15.34 | 568 | Serine-rich AF-9 protein. New # |
| A_Geneseq_36:W23793 | - | 56.50 | 106.49 | 20.40 | 917 | AS-30D tumour type II hexokinas |
| A_Geneseq_36:W37437 | - | 56.50 | 106.49 | 20.40 | 917 | Rat hexokinase II. Mammalian ce |
| A_Geneseq_36:W37429 | - | 56.50 | 106.49 | 20.40 | 917 | Rat hexokinase II. Mammalian ce |
| A_Geneseq_36:W90028 | + | 55.00 | 116.94 | 22.68 | 216 | Expressed antigen for clone Z3C |
| A_Geneseq_36:W90027 | + | 55.00 | 116.32 | 23.06 | 230 | Expressed antigen for clone Z3C |
| A_Geneseq_36:W20566 | + | 55.00 | 116.19 | 23.13 | 233 | Helicobacter pylori outer membr |
| A_Geneseq_36:W24691 | + | 55.00 | 116.19 | 23.13 | 233 | Helicobacter pylori outer membr |
| A_Geneseq_36:Y11004 | + | 55.00 | 115.70 | 23.44 | 245 | H. pylori ORF hp4el3394_5908553 |
| A_Geneseq_36:W20997 | + | 55.00 | 115.54 | 23.54 | 249 | H. pylori ORF outer membrane protel |
| A_Geneseq_36:W30648 | - | 55.50 | 109.96 | 24.01 | 499 | A. thaliana xylan synthase prot |
| A_Geneseq_36:P94635 | - | 55.50 | 107.40 | 25.70 | 648 | B. thuringiensis alpha-amylase |
| A_Geneseq_36:W41819 | - | 55.50 | 107.32 | 25.75 | 653 | H. annuus trehalose phosphate s |
| A_Geneseq_36:W30641 | - | 55.50 | 107.03 | 25.95 | 673 | Arabidopsis thaliana xylan synt |
| A_Geneseq_36:W25719 | + | 55.50 | 106.84 | 26.08 | 686 | Human alpha meltrin protein fir |
| A_Geneseq_36:W00097 | - | 55.00 | 111.28 | 26.34 | 384 | cAMP phosphodiesterase encoded |
| A_Geneseq_36:R27511 | - | 55.00 | 110.91 | 26.61 | 399 | Human R3BP. Receptors for bombe |
| A_Geneseq_36:W41813 | - | 55.50 | 105.59 | 26.96 | 779 | H. annuus trehalose phosphate p |
| A_Geneseq_36:R04749 | - | 54.50 | 114.86 | 27.23 | 235 | Amino acid sequence of maize ze |
| A_Geneseq_36:R04751 | - | 54.50 | 114.86 | 27.23 | 235 | Amino acid sequence of maize ze |
| A_Geneseq_36:R56696 | - | 55.00 | 106.37 | 30.00 | 633 | CryIIA insecticidal crystal prd |
| A_Geneseq_36:W75772 | - | 55.00 | 106.34 | 30.02 | 635 | Amino acid sequence of lepidote |
| A_Geneseq_36:W84272 | + | 56.50 | 91.79 | 30.10 | 4096 | A DNA-dependent protein kinas |
| A_Geneseq_36:R63217 | - | 54.50 | 110.26 | 30.74 | 375 | Human alpha-2,3-sialyltransfer |
| A_Geneseq_36:R62808 | - | 54.50 | 110.26 | 30.74 | 375 | Alpha 2, 3-sialyl transferase. |

| | | | | | | |
|-------------------------------|---|----------------------------|--------|-------|-----|------------------------------|
| A_Geneseq_36:R65242 | - | 54.50 | 110.26 | 30.74 | 375 | Human ST3N sialyltransferase |
| A_Geneseq_36:W55518 | - | 54.00 | 112.28 | 33.12 | 269 | H. pylori ORF 07ep11916_5913 |
| A_Geneseq_36:W00641 | - | 54.00 | 111.47 | 33.83 | 292 | Infectious laryngotracheitis |
| seq_name: A_Geneseq_36:W75057 | | | | | | |
| seq_documentation_block: | | | | | | |
| ID | W75057 | standard; Protein; 467 AA. | | | | |
| AC | W75057; | | | | | |
| DT | 28-JAN-1999 | (first entry) | | | | |
| DE | Human secreted protein encoded by gene 1 clone HGCM20. | | | | | |
| KW | Human; secreted protein; fusion protein; gene therapy; protein therapy; | | | | | |
| KW | diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; | | | | | |
| KW | developmental abnormality; foetal deficiency; blood; allergy; renal; | | | | | |
| KW | immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; | | | | | |
| KW | inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS; | | | | | |
| KW | cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; | | | | | |
| KW | osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; | | | | | |
| KW | endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm. | | | | | |
| OS | Homo sapiens. | | | | | |
| FH | Key | Location/Qualifiers | | | | |
| FT | Misc-difference 97 | /label= unknown | | | | |
| FT | Misc-difference 119 | /label= unknown | | | | |
| FT | Misc-difference 240 | /label= unknown | | | | |
| PN | W09839446-A2. | | | | | |
| PD | 11-SEP-1998. | | | | | |
| PE | 06-MAR-1998; | U04492. | | | | |
| PR | 07-MAR-1997; | US-038621. | | | | |
| PR | 07-MAR-1997; | US-040161. | | | | |
| PR | 07-MAR-1997; | US-040162. | | | | |
| PR | 07-MAR-1997; | US-040163. | | | | |
| PR | 07-MAR-1997; | US-040333. | | | | |
| PR | 07-MAR-1997; | US-040334. | | | | |
| PR | 07-MAR-1997; | US-040336. | | | | |
| PR | 07-MAR-1997; | US-040626. | | | | |
| PR | 11-APR-1997; | US-043311. | | | | |
| PR | 11-APR-1997; | US-043312. | | | | |
| PR | 11-APR-1997; | US-043313. | | | | |
| PR | 11-APR-1997; | US-043314. | | | | |
| PR | 11-APR-1997; | US-043315. | | | | |
| PR | 11-APR-1997; | US-043568. | | | | |
| PR | 11-APR-1997; | US-043569. | | | | |
| PR | 11-APR-1997; | US-043576. | | | | |
| PR | 11-APR-1997; | US-043578. | | | | |
| PR | 11-APR-1997; | US-043580. | | | | |
| PR | 11-APR-1997; | US-043669. | | | | |
| PR | 11-APR-1997; | US-043670. | | | | |
| PR | 11-APR-1997; | US-043671. | | | | |
| PR | 11-APR-1997; | US-043672. | | | | |
| PR | 11-APR-1997; | US-043674. | | | | |
| PR | 23-MAY-1997; | US-047492. | | | | |
| PR | 23-MAY-1997; | US-047500. | | | | |
| PR | 23-MAY-1997; | US-047501. | | | | |
| PR | 23-MAY-1997; | US-047502. | | | | |
| PR | 23-MAY-1997; | US-047503. | | | | |
| PR | 23-MAY-1997; | US-047581. | | | | |
| PR | 23-MAY-1997; | US-047582. | | | | |
| PR | 23-MAY-1997; | US-047583. | | | | |
| PR | 23-MAY-1997; | US-047584. | | | | |
| PR | 23-MAY-1997; | US-047585. | | | | |
| PR | 23-MAY-1997; | US-047586. | | | | |
| PR | 23-MAY-1997; | US-047587. | | | | |
| PR | 23-MAY-1997; | US-047588. | | | | |
| PR | 23-MAY-1997; | US-047589. | | | | |
| PR | 23-MAY-1997; | US-047590. | | | | |
| PR | 23-MAY-1997; | US-047592. | | | | |
| PR | 23-MAY-1997; | US-047593. | | | | |
| PR | 23-MAY-1997; | US-047594. | | | | |
| PR | 23-MAY-1997; | US-047595. | | | | |
| PR | 23-MAY-1997; | US-047596. | | | | |

PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057761.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-609887/51.
DR N-PSDB; V34154.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 278-280; 447pp; English.
CC This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line.
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. V34145) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 70 novel genes and their fragments (nucleic acid
CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC *Specific uses are described for each of the 70 polynucleotides, based on

CC which tissues they are most highly expressed in (see V34154 for described
CC uses).
SQ Sequence 467 AA:

alignment_scores:
Quality: 71.00 Length: 75
Ratio: 1.919 Gaps: 3
Percent Similarity: 49.333 Percent Identity: 30.667

alignment_block:
US-09-030-606-224/rev x W75057 ..

Align seg 1/1 to: W75057 from: 1 to: 467

229 TATCAAGCACCTTTAAGGGCTTCCAT..... 203
||| ||| |||::: |||||
180 TyrTyrSerLeuPheGluTrpPheHisProLeuPheLeuGluAspGluSe 196
202TTYGAGAAAGTAGTATT.....TCCTCTCTCTC 172
|||:::||| ||| |||:::|||
196 rSerSerPheHisLysArgGlnPheProValSerLysThrLeuProGluL 213
171 TTTTAGACATATGATAAATAATATGCGCTTATCCTTTAACCATGTATC 122
|| ||| :::: ||| ||| |||
213 eUTyrgLuleValAsnAsnTyrgLlnProGluValLeu..... 225
121 CATTTTTTTGTGATGGGAGTAAGGGGTACACACTCCCTACTACATGT 72
::: |||||::: |||::: |||:::|||
226TrpSeraspGlyAspGlyGlyAlaProaspGlnTyrrTpasn** 240
71 CACAACTTTTCTTTTGTGTTC 47
|||:::|||::: |||:::|
240 *ThrGlyPheLeuAlaTrpLeuTyr 248

seq_name: A_Geneseq_36:Y13378

seq_documentation_block:
ID Y13378 standard; Protein; 467 AA.
AC Y13378;
DT 25-JUN-1999 (first entry)
DE Amino acid sequence of protein PRO260.
KW Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW anti-thrombotic; wound healing; tissue repair.
OS Homo sapiens.
PN WO9914328-A2.
PD 25-MAR-1999.
PF 16-SEP-1998; U19330.
PR 25-NOV-1997; US-066840.
PR 17-SEP-1997; US-059113.
PR 17-SEP-1997; US-059115.
PR 17-SEP-1997; US-059117.
PR 17-SEP-1997; US-059119.
PR 17-SEP-1997; US-059121.
PR 17-SEP-1997; US-059122.
PR 17-SEP-1997; US-059184.
PR 18-SEP-1997; US-059263.
PR 18-SEP-1997; US-059266.
PR 15-OCT-1997; US-062125.
PR 17-OCT-1997; US-062285.
PR 17-OCT-1997; US-062287.
PR 21-OCT-1997; US-063486.
PR 24-OCT-1997; US-062814.
PR 24-OCT-1997; US-062816.
PR 24-OCT-1997; US-063045.
PR 24-OCT-1997; US-063120.
PR 24-OCT-1997; US-063121.
PR 24-OCT-1997; US-063127.

PR 24-OCT-1997; US-063128.
PR 27-OCT-1997; US-063329.
PR 27-OCT-1997; US-063327.
PR 28-OCT-1997; US-063541.
PR 28-OCT-1997; US-063542.
PR 28-OCT-1997; US-063544.
PR 28-OCT-1997; US-063549.
PR 28-OCT-1997; US-063550.
PR 28-OCT-1997; US-063564.
PR 29-OCT-1997; US-063435.
PR 29-OCT-1997; US-063704.
PR 29-OCT-1997; US-063732.
PR 29-OCT-1997; US-063738.
PR 29-OCT-1997; US-063734.
PR 29-OCT-1997; US-064215.
PR 31-OCT-1997; US-063735.
PR 31-OCT-1997; US-063870.
PR 31-OCT-1997; US-064103.
PR 03-NOV-1997; US-064248.
PR 07-NOV-1997; US-064809.
PR 12-NOV-1997; US-065186.
PR 17-NOV-1997; US-065846.
PR 18-NOV-1997; US-065693.
PR 21-NOV-1997; US-066120.
PR 21-NOV-1997; US-066364.
PR 24-NOV-1997; US-066772.
PR 24-NOV-1997; US-066466.
PR 24-NOV-1997; US-066770.
PR 24-NOV-1997; US-066511.
PR 24-NOV-1997; US-066453.
PA (GETH) GENENTECH INC.
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
DR WPI; 99-229533/19.
DR N-PSDB; X52249.
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
PS Claim 12; Fig 72; 320pp: English.
CC Y13344-403 represent and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to growth
CC or survival of nerve cells including Parkinson's disease, Alzheimer's
CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
SQ Sequence 467 AA;

alignment_scores:

Quality: 70.00 Length: 75
Ratio: 1.892 Gaps: 3
Percent Similarity: 49.333 Percent Identity: 30.667

alignment_block:

US-09-030-606-224/rev x Y13378 ..

Align seg 1/1 to: Y13378 from: 1 to: 467

229 TATCAAGCACCTTTAAGGCTTCAT..... 203
||| ||| |||::: |||||

180 TyrTyrSerLeuPheGluTrpPheHisProLeuPheLeuGluAspGluSe 196
202TTYGAAAGTAGTATT.....TCTCTCTCTCTC 172
196 rSerSerPheHisLysArgGlnPheProValSerLysThrLeuProGluL 213
171 TTTAGACATATGATAAATATTCGCCCTTATCCTTAACCATGATC 122
213 eUTyrgLUleuValAsnAsnTYrgLInProGluValLeu..... 225
121 CATTTTTTTGTGATGGAGTAGGGGTACACACTCCCTACTACATGT 72
226TrpSerAspGlyAspGlyGlyAlaProAspGlnTYrTrpAsnSe 240
71 CACAACTTTTCTTTTGTGTTTC 47
240 rThrgLUleuAlaTrpLeuTYr 248
|||:::||||: |||:::
|||:::||||: |||:::

seq_name: A_Geneseq_36:R88732

seq_documentation_block:

ID R88732 standard; Protein; 663 AA.
AC R88732.
DT 17-SEP-1996 (first entry)
DE S.aureus topoisomerase IV grIB subunit.
KW GrIA; GrIB; Staphylococcus aureus; topoisomerase IV; type II; E.coli;
KW DNA gyrase; PCR; amplification; primer; polymerase chain reaction;
KW bacterium; Bacillus subtilis; Gyra; GyrB; ParC; Pare; operon; Haloferax;
KW termination codon; initiation; frameshift; translation; fluoroquinolone;
KW antibacterial; inhibition; relaxation; decatenation.
OS Staphylococcus aureus.
PN WO9603516-A1.
PD 08-FEB-1996.
PF 26-JUL-1995; F01001.
PR 27-JUL-1994; FR-009288.
PA (RHON) RHONE POULENC RORER SA.
PI Blanche F, Cameron B, Crouzet J, Famechon A, Ferrero L;
DR WPI; 96-117059/12.
DR N-PSDB; T12563.
PT Staphylococcus aureus topo:isomerase 4 sub-unit coding sequence -
PT useful e.g. to screen for potential antibacterials
PS Claim 12; Page 39-41; 60pp: French.
CC This is the amino acid sequence of the grIB subunit of the Staphylococcus
CC aureus topoisomerase IV, a member of the type II topoisomerases. The
CC other extensively studied member of the family is E.coli DNA gyrase.
CC Topoisomerase IV comprises 2 subunits - GrIA (R88731) and GrIB. GrIB is
CC a protein which has a pI of 7.21 and mol. wt. ca. 74.3 kD. The corresp.
CC gene was isolated by PCR amplification using primers T12566-7 based on
CC sequence similarity with other bacterial topoisomerase IV genes e.g. the
CC E.coli and B.subtilis GyrB, or E.coli Pare genes. The 1 kb fragment
CC amplified, corresp. to residues 98-430 of E.coli GyrB, was used to probe
CC a S.aureus genomic DNA library to obtain the grIB operon sequence. In
CC the operon, the grIB gene termination codon straddles the grIA initiation
CC codon, implying that a frameshift is required during translation to
CC accurately translate the grIA mRNA. A similar linkage is seen in the
CC gyra and B genes in Haloferax sp. Topoisomerase IV is a target for the
CC fluoroquinolone antibacterials. The protein can thus be used to screen
CC for new antibacterials and to screen for cpds. Inhibiting ATP-dependent
CC DNA relaxation and DNA decatenation.
SQ Sequence 663 AA;

alignment_scores:

Quality: 66.00 Length: 72
Ratio: 1.650 Gaps: 3
Percent Similarity: 55.556 Percent Identity: 30.556

alignment_block:

US-09-030-606-224 x R88732 ..

Align seg 1/1 to: R88732 from: 1 to: 663

43 ATAGGACACACAAAGAAAGATTGTGACATGTGTAGTAGGAGTGTGT 92


```

|||||:||||| :|||||:||||| ||| ||| ||| |||
29 IleglySerThrAspLysArgGlyLeuHisHisLeuValTyrGluIleVal 45

93 ACCCCTTACTCCCATCAAAAAAATGGATACATGGTTAAAGATARA 142
  | :| :| :| |||||
45 LaspasnservAlaspGluValIleuasnglyTyr..... 56

143 AGGCATATTTTATCATATGTTCTAAAGAGAGAAGAGAGAAAATACTA 192
  ||||| :|:|:|:| :|:|:|:| :|:|:|
57 ..GlyAsngluIleAspValThrIle..AsnLysaspGlySerIleSer 71

193 CTTTCTCRAAATGGAAGCCCTTAAGGTGCTTTGATACTGGAAGACACA 242
  :|:|:| :| ||||| ||| :|:| ||| |||||
72 Ilegluaspasngly.....ArgGlyMetProThrGlyIleHisIstly 85

243 ATGTGCCCGTCCATCC 258
  | |||:|:|:|:|:|
85 sSerGlyLysProThr 90

seq_name: A_Geneseq_36:y12066

```

seq_documentation_block:
ID Y12066 standard; Protein; 100 AA.
AC Y12066;
DT 18-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO: 379.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN W03906554-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1238.
PR 01-AUG-1997; US-905134.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153784/13.
DR N-PSDB; X40899.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
PS C1a1m 34; Page 500; 622pp; English.
CC X40826 to X41093 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y01602 and
CC Y11994 to Y12260, respectively. The proteins given represent the signal
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy.
CC The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.
SQ Sequence 100 AA;

```

alignment_scores:      Quality: 63.50      Length: 63
                       Ratio: 2.352      Gaps: 3
Percent Similarity: 42.857      Percent Identity: 26.984

alignment_block:
US-09-030-606-224/rev x Y12066      ..

```

```

Align seg 1/1 to: Y12066 from: 1 to: 100

135 TTTAACCATGATCCATTTTTTTTTTGATGGGAGTAAGGG..... 94
    |||::|||::|||::|||::|||::|||::|||::|||::|||
11 PheSerHisAlaSerLeuPheMetValGlySerThrGlySerLeu1 27
93 .....TACACACTCCCTACTACA..... 76
    |||||
27 eleuLeuThrSerCysPheYrThrLeuValSerSerThrPheLeu1L 44
75 .....ATG 73
    |||
44 yslSerSerLeuLeuLeuLeuPheThrGluThrSerValLeuMet 60
72 TCACAACATTT.....TCCTTTTGTTGTTCCATTGG 40
    ::|||::|||::|||::|||::|||::|||::|||
61 leuLysThrPheValAlaAsnSerCysCys**YeuTirp 73

```

```

seq_name: A_Geneseq_36:R77168
seq_documentation_block:
ID   R77168 standard; Protein; 202 AA.
AC   R77168;
DT   04-DEC-1995 (first entry)
DE   Arabidopsis condensin enzyme clone CE17.
KW   Arabidopsis; condensin enzyme.
OS   Arabidopsis thaliana.
PN   W09515387-A.
PD   08-JUN-1995.
PE   30-NOV-1994; U13686.
PF   30-NOV-1993; US-160602.
PR   23-JUN-1994; US-265047.
PI   (CALJ ) CALGENE INC.
PI   lardizabal KD, Lassner MW, Metz JG;
DR   WPI; 95-215267/28.
DR   N-PSDB; Q90214.
PT   Production of very long chain fatty acid(s) in plant(s) - to produce
PT   drought and stress resistant transgenic plant(s)
PS   Example: Figure 9; 149pp; English.
CC   The CE15 and CE20 Brassica CDNA sequences (see Q90210,Q90211
CC   and Q90212) and the condensin enzyme encoding sequence from
CC   fojoba (Q90208) were used in determining primers Q90221-
CC   Q90225 from conserved AAs. These primers were variously used to
CC   PCR (RTPCR) amplify fragments from RNA isolated from developing
CC   seeds of Lunaria annua, Tropaeolu majus (Nasturtium), and
CC   green illiques of Arabidopsis thaliana. The primers most
CC   successfully utilised were Q90221 and Q90222. These primers were
CC   used to produce three clones encoding a portion of the elongase
CC   condensin enzyme from Arabidopsis, designated ARAB CE15, ARAB CE17
CC   and ARAB CE19 (see Q90213/R77167; Q90214/R77168; Q90215/R77169
CC   respectively. Clones were also identified from Lunaria and
CC   Nasturtium.
SQ   Sequence 202 AA;

```

```

alignment_scores:
    Quality: 63.50      Length: 68
    Ratio: 1.477      Gaps: 5
Percent Similarity: 63.235      Percent Identity: 33.824

alignment_block:
US-09-030-606-224/rev x R77168      ..

Align seg 1/1 to: R77168 from: 1 to: 202

205 CATTYGAAGAAGTAGTATTCTCTCTCTCTTTTGAACATATGAT 156
    ||||| |||:||||| ||||| ::| :|||
12 HisphenhelysleuemetPhe..LeuProleumetalavalLeupheme 27

155 AAAATATTGCCCTTATATCCCTTTAACCATGTATCCATTTTTTTTGATG 106
    : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 t.AsnValSerleuSeuserleuAsnHisleugin..... 38

```



```

105 GGGAGTAGGGGTACACACTCCCTACTACAATGTCAAACTTTTCTT 56
      ||| ||||| |||:||||:||||
39 .....Leu..TyrTyrAsnSerThrGlyPheIlePhe 48
55 TTGTGTTCCTATTGGTTGTA..ACTGTACTATTTTCTAACAAGAACC 9
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
49 ValIleThrLeuAlaIleValGlySerIleValPheMetSerArgPr 65
      8 T 8
      |
65 O 65

```

seq_name: A_Geneseq_36:Y07052

seq_documentation_block:
ID Y07052 standard; Protein; 312 AA.
AC Y07052;
DT 02-JUL-1999 (first entry)
DE Renal cancer associated antigen precursor sequence.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
PN MO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI Pfeundschnuh M, Sahln U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI: 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Disclosure; Page 439-440; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 312 AA;

```
alignment_scores:
    Quality: 62.50
    Ratio: 2.155
    Percent Similarity: 46.032
    Length: 63
    Gaps: 3
    Percent Identity: 34.921
```

alignment_block:

US-09-030-606-224 X Y07052

Align seg 1/1 to: Y07052 from: 1 to: 312

```

115 AAAATGCATACATGCTTAAAGCATARAAGGGC..... 147
    |||::|||::|||::|||  |||||
186 LysGlnGlyTyrMetMetLys....LysGlyHisArgArgLysAsnTr 200

```

```

148 .....AATATTTATCATATGTT... 165
200 pThrgluarGtrPheValLeuLysProAsnIleIleSerTyrTyrValS 217
166 .....CTAAAGAGAGAAGAGAAATACTACTTCTCRAAATGGA 207
      |||||:::||||: |||||::: |||
217 erGIuAspLeuLysAspLysIcylAspIleLeuLeuAspGIuAsnCys 233
208 AGCCCTTAAGGTGCTTGATACTGAAAGACACAATGT 246
      ||| ::||| |||||
234 CysValGIuSerLeuProAspLysAspGIyLysLysCys 246

```

seq_name: A_Geneseq_36:W97757

```
seq_documentation_block:
ID      W97757 standard; Protein; 585 AA.
AC      W97757;
DT      21-MAY-1999 (first entry)
DE      S-region transfer activity-70 (SRTA-70).
KW      SRTA-70; S-region transfer activity-70; recombination protein;
KM      switch recombination; mouse; immunomodulator; cancer; allergy;
KV      therapy.
OS      Mus sp.
FH      Key          location/Qualifiers
FT      Peptide     222..225
FT              /note="nuclear localisation signal"
FT      Peptide     291..294
FT              /note="nuclear localisation signal"
FT      Peptide     360..376
FT              /note="nuclear localisation signal"
FT      MO9903991-A1.
PD      28-JAN-1999.
PF      16-JUL-1998; IB1191.
PR      18-JUL-1997; EP-112326.
PA      (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI      (REGC ) UNIV CALIFORNIA.
DR      Jessberger R, Wabl M;
        MPI; 99-132235/11.
N-PSDB: X07259.
PT      New DNA encoding S-region transfer activity-70 protein - used to
PT      identify specific binding agents and for modulating the immune
PT      response e.g. in cases of cancer and allergy
PS      Claim 8; Fig 4; 37pp; English.
CC      This polypeptide comprises murine SRTA-70, or S-region transfer
CC      activity-70, a 70 kDa protein that mediates switch recombination.
CC      The amino acid sequence was deduced from a cDNA clone (see X07259)
CC      isolated from a mouse spleen library. The invention relates to the
CC      isolation, purification and characterisation of proteins mediating
CC      switch recombination. It also relates to recombinant SRTA-70
CC      proteins, vectors and host cells. SRTA-70 polypeptides are used:
CC      (1) for modulating the immune response; (ii) for identifying
CC      specific (antagonists); (iii) to identify proteins (and their
CC      genes) involved in class-switch recombination or other B cell
CC      functions, and (iv) to raise antibodies.
SQ      Sequence   585 AA;
```

```

alignment_scores:
  Quality: 62.50
  Ratio: 2.155
  Percent Similarity: 46.032
  Length: 63
  Gaps: 3
  Percent Identity: 34.921

```

alignment_block:

US-09-030-606-224 X W97757

Align seg 1/1 to: W97757 from: 1 to: 585

```

115 AAAATGATACATGGTTAAAGCATARAAGGC..... 147
    |||:::|||||||:::|||    |||||
213 LysGlnGlyTyrMetMetLys.....LysGlyHisLysArgLysAsnTr 227
    |||:::|||||||:::|||    |||||
148 .....AATATTTATCATATGTT..... 165
    |||:::|||||||

```


227 pThrcIuArgTrpHeValleuLysProAsnIleIleSerTyrTyrValS 244
166CTAAAGAGAGAGAGAGAAATACTACTTCTCRAAATGA 207
244 ercIuAspleuLysAspLysLysGlyAspIleLeuLeuAspGluAsnCys 260
208 AGCCCTTAAGGTGCTTGATCTACTGAAGACACAAATGT 246
261 CysValGluSerLeuProAspLysAspGlyLysLysCys 273

seq_name: A_Geneseq_36:R20030

seq_documentation_block:

ID R20030 standard; Protein; 633 AA.
AC R20030;
DT 07-APR-1992 (first entry)
DE Bacillus thuringiensis CryIIb crystal toxin.
KW B.t. crystal protein; insecticide; Lepidoptera; larva; corn earworm;
OS Heliothis zea.
PN US5073632-A.
PD 17-DEC-1991.
PF 11-JUL-1989; 379015.
PR 16-APR-1987; US-039542.
PR 11-JUL-1989; US-379015.
PA (ECOG-) ECOGEN INC.
PI Donovan WP;
DR WPI; 92-016224/02.
DR N-PSDB; Q20202.
PT Bacillus thuringiensis cryIIb protein gene - used to obtain the
PT protein for use as an insecticidal cpd. against lepidopteran
PT insects
PS Disclosure; Fig 6; 30pp; English.
CC The cryIIb gene was isolated from a B.t. strain using the cryIIa
CC gene as probe. The cryIIb gene encodes a 633 amino acid protein of
CC calculated mol. wt. 70,749. The insecticidal toxin CryIIb is twice
CC as toxic as the CryIIa protein against the lepidopteran Heliothis
CC zea.
SQ Sequence 633 AA;

alignment_scores:
Quality: 59.00 Length: 81
Ratio: 1.311 Gaps: 5
Percent Similarity: 55.556 Percent Identity: 35.802

alignment_block:

US-09-030-606-224/rev x R20030 ..

Align seg 1/1 to: R20030 from: 1 to: 633

241 TGTGTCCTTCAGTATCAAGCACCTTTAAGGCTTCCATTYTGAGAAAGT 192
223 CysIleAsnThrTyrGlnSerAlaPheLysGlyLeuAsnThrArg.... 237
191 AGTATTTTCTCTCTCTCTTTAGAACATATGATAAATATTGCCCTT 142
238LeuHisAspMetLeuGluPheArgThr.TyrMetPheLeu 250
141 YTATCCTTTAACCAAGTATCCATT.....TTTTTT.....TT 110
251 AsnValPheGluTyrValSerIleTrpSerPhePheLysTyrGlnSerLe 267
109 GATGGGAGTAAGGGGTACACACTCCCTACTACATGTCA.....C 69
267 uIeuValSerSerGlyAlaAsnLeuTyrAlaSerGlySerGlyProGlnG 284
68 AAACCTTTTCTTTTGTGTCCTATTTGTTGTAACGTAC 28
284 InThrGlnSerPheThrSerGlnAspTrpPheLeuTyr 297

seq_name: A_Geneseq_36:W88448

seq_documentation_block:

ID W88448 standard; Protein; 1296 AA.
AC W88448;
DT 26-APR-1999 (first entry)
DE Caenorhabditis elegans NPc1 protein orthologue.
KW Niemann-Pick disease type C; NPc1 gene; worm; orthologue;
KW diagnosis; therapy; animal model; cholesterol; neurodegeneration.
OS Caenorhabditis elegans.
PN W09901555-A1.
PD 14-JAN-1999.
PF 02-JUL-1998; U13862.
PR 03-JUL-1997; US-051682.
PA (USSH) US DEPT HEALTH & HUMAN RESOURCES.
PI Carstea ED, Gu J, Loftus SK, Morris JA, Pavan WJ,
PI Pentchev PG, Rosenfeld MA, Tagle DA,
DR WPI; 99-106056/09.
DR N-PSDB; X06876-77.
PT New isolated gene, NPC-1 - is associated with Niemann-Pick type C
PT disease, used to develop products for the study, diagnosis and
PT therapy of the disease
PS Disclosure; Page 91-95; 101pp; English.
CC This polypeptide comprises the Caenorhabditis elegans orthologue
CC of the human NPc1 polypeptide that is associated with
CC Niemann-Pick disease type 2 (NP-C). The polypeptide shows
CC extensive identity (30%) and similarity (55%) to the human
CC NPc1 protein (see W88445). Biochemical and genetic analysis of
CC yeast, worm and murine NPc1 model systems will provide resources
CC for understanding the role of NPc1 in intracellular cholesterol
CC homeostasis and in the aetiology of neurodegeneration in NP-C
CC disease. The provision of the human NPc1 cDNA sequence (see
CC X06873) enables methods of detecting the presence of mutations in
CC the NPc1 gene, and thereby facilitates the determination of
CC whether an individual is an NP-C sufferer or carrier.
SQ Sequence 1296 AA;

alignment_scores:
Quality: 59.50 Length: 114
Ratio: 1.062 Gaps: 7
Percent Similarity: 49.123 Percent Identity: 22.807

alignment_block:

US-09-030-606-224/rev x W88448 ..

Align seg 1/1 to: W88448 from: 1 to: 1296

278 AGTCATGCACTYTAAGGAGATGA.....CGGCCA..... 246
1153 SerHisIleSerAsnValGluGlyIleLeuAsnArgProSerLeuLe 1169
245CATTTGTCCTTCAGTATCAAGCACCTTTAAGGCTTC 206
1169 uAspAlaSerHisIleLeuAspProLeuLeuLysAlaGluGlyIle. 1185
205 CATTTYGAGAAAGTATTTCTCTCTCTCTTTAGAACATATGAT 156
1186AspLysAlaIleGlyArgAspPheLysTyrHisAsnArg 1199
155 AAAATATGCCCCTYTAFCCTTA.....ACCATGTAT..... 123
1200 SerLeuLeuProIleTyrAlaIlePheThrAlaMetTyrLeuProAsnAl 1216
122CAATTTT..... 114
1216 alysserSerTyrArgThrGlySerProIlePheValAspThrAspLeuL 1233
113 ..TTTGATGGGAGTAAGGGTACACAC.....TCCCTA 81
1233 eupheThrTyrSerIleProTyrIleHisPheValIleSerValSerPhe 1249
80 CTACAATGTACAAACTTTTCTTTTGTGTCCTATTTGCT 39
1250 IleLysCysHisValLeuTyrIleTyrAlaAlaThrIleGly 1263

PN W09738136-A1.
PD 16-OCT-1997.
PF 09-APR-1997; U05710.
PR 06-MAR-1997; US-812481.
PR 09-APR-1996; US-015083.
PR 09-APR-1996; US-630048.
PR 06-MAR-1997; US-038614.
PA (BARD-) BARD DIAGNOSTIC SCI INC.
PI Enfield DL, Hass GM, Kinders RJ;
DR WPI: 97-512742/47.
DR N-PSDB: V02792.
PT Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement factor
PT H related antigen, or nucleic acid encoding it
PS Example 6B; Fig 7B; 104pp; English.
CC This partial protein sequence represents a region of the human
CC tumour-associated complement factor H (CFH). This sequence is used in the
CC identification of complement factor H related proteins and antigens from
CC clone PRBS3FH2910 (see W39157-W39159). The detection of this protein and
CC a CFH antigen can be used in screening or for the treatment of renal or
CC urogenital cancer, e.g. bladder, cervical or prostate cancer. Agents that
CC may modulate this antigen could be used in the manufacture of a
CC medicament for the treatment of a tumour cell.
SQ Sequence 290 AA;

alignment_scores: Quality: 57.50 Length: 90
 Ratio: 1.223 Gaps: 5
Percent Similarity: 52.222 Percent Identity: 27.778

alignment_block:
US-09-030-606-224 x W39156 ..

Align seg 1/1 to: W39156 from: 1 to: 290

82 AGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATACATGGT 131
 |||::: ::::||| ::|||::|||::|||
141 ArgAspThrSerCysValAsnProProThrValGlnAsnAlaTyrIleVa 157
132 T.....AAAGATARAAGGCAATATTTATCATATGTTTC 166
 | ||| ::|||::: ::|||
157 lSerArgGlnMetSerIleTyrProSerGlyGluArgValArgTyrGlnC 174
167 TAAAAGAG.....AAGAAGAGAAAATACTACTTCT 198
 ::::: ||||| ::|||::: |||:::
174 ysArgSerProTyrGlnMetPheGlyAspGluValMetCysLeuAsn 190
199 CRAATGGAAGC...CCTTAAAGGTGCTTTGATGACTGAAGACACAAATG 245
 ||| ::: ||| ::||| |||::: |||
191 GlyAsnTrpThrGluProGlnCysLysAspSerThrGly...LysCys 206
246 TGGCCGTCCATCCTCTTARAGTTGCATGACTGACACGTAACGT 295
 |||| ||| |||||::: |||::: :::
206 sGlyProPro.ProProIleAspAsnGlyAspIleThrSerPheProLeu 222
296 GCAGTTTARACTCMGCA 313
 ::||| ::|||
223 SerValTyrAlaProAla 228

seq_name: A_Geneseq_36:R56285

seq_documentation_block:

ID R56285 standard; Protein: 461 AA.
AC R56285;
DT 03-MAR-1995 (first entry)
DE Sequence of human alpha-L-fucosidase including signal peptide.
KW Anti-adhesion enzyme; alpha-L-fucosidase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..22
FT /label= signal
FT protein 23..461

FT W09414472-A. /label= mature
PN 07-JUL-1994.
PD 21-DEC-1993; U12464.
PF 22-DEC-1992; US-994650.
PR (REGC) UNIV CALIFORNIA.
PI Carson DA, Wicks I;
DR WPI: 94-234351/28.
DR N-PSDB: Q66537.
PT Inhibiting adhesion of cells to selectin-contg. receptors - by
PT treatment with enzyme degrading carbohydrate residues in selectin
PT adhesion ligands, partic. for treating inflammation
PS Disclosure; Page 27-30; 42pp; English.
CC The AA sequence in R56285 is derived from sequence information obtd.
CC through isolation of several human alpha-L-fucosidase cDNA clones
CC reported by Occhiodoro et al. at Biochem. Biophys. Res. Comm.,
CC 164:439-445 (1989). The full-length nucleotide sequence for alpha-L-
CC fucosidase is available from the GENBANK molecular sequence database
CC under Accession No. M28099. The DNA sequence in Q66537 is described
CC as genomic.
SQ Sequence 461 AA;

alignment_scores: Quality: 57.50 Length: 64
 Ratio: 1.597 Gaps: 2
Percent Similarity: 56.250 Percent Identity: 28.125

alignment_block:
US-09-030-606-224/rev x R56285 ..

Align seg 1/1 to: R56285 from: 1 to: 461

238 GTCCTTCAGTATCAAGCACCTTTAAGGCTTCCATTTTGAGAAAGTAGT 189
 ::|||::: ::::: ||||| ||||| :::::
189 LeuLeuAspLysLysAsnGlyPheLysThrGlnHisPheValSerAlaL 205
188 ATTTCTCTCTCTCTCTTTAGAACATATGATAAATATTCGCCCTTTA 139
 :::::|||::| ::: ::||| |||
205 s...ThrMetProGlnLeuTyrAspLeuValAsnSerTyrLysProAspL 221
138 TCCTTTAACCATGATCCATTTTGTGATGGGAGTAAGGGGTACAC 89
 ::::: ::|||
221 eulle.....TipSerAspGlyGluTrpGluCysPro 231
88 ACTCCCTACTACATGTCACAACCTTTTCTTTTGTGTGTC 47
 |||::||| ||||| |||:::
232 AspThrTyrTrpAsnSerThrAsnPhenLeuSerTrpLeuTyr 245

•
•
•


```
seq_documentation_block:
; Sequence 30, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaanl, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; TITLE OF INVENTION: All-1 Region
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-320-559-30
```

```
alignment_scores:
Quality: 57.00 Length: 98
Ratio: 1.462 Gaps: 3
Percent Similarity: 39.796 Percent Identity: 23.469
```

alignment_block:

US-09-030-606-224/rev x US-08-320-559-30 ..

Align seg 1/1 to: US-08-320-559-30 from: 1 to: 568

```
250 GGCACATTGTGTCCTCAGTATCAAGACCTTTAAGGCTTC..... 206
||||| : : : : : ||| : : |||||
14 GlyHisArgAlaGlnValArgLysLysProThrValGlnGlyPheThrH1 30
205 .....CATT 202
|||||
30 sAspTrpMetValPheValArgGlyProGlnHisSerAsnIleGlnHisP 47
```

```
201 TY...GAGAAAGTAGTATTCTCTCTCTCTCTTTAGACATATGATA 155
|| ||||| ||||| ||| : : : : :
47 heValGlnLysValAlaPheHisIleHisGluSerPheProArgProLys 63
154 AATATATGCCCTTATATCCTTTAACCATGTATCCATTTTGTGATGG 105
: : : |||
64 ArgValCys..... 66
104 GGAGTAAGGGGTACACACTCCCTACTACATGTCAACAACCTTTCTTTT 55
||| ||| : : : : :
67 .....LysAspProProTyrLysValGlnGluSerGlyTyrA 79
54 TGTGTTCTTATGTTGTTACTGTACTATTCTTACACAGAAG 11
||| : : : ||| ||| : : ||||| : :
79 IagIyPheIleLeuProIleGlnValTyrPheLysAsnLysGln 93
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-04496-30
```

```
seq_documentation_block:
; Sequence 30, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaanl, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewitz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1242
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-04496-30
```

```
alignment_scores:
Quality: 57.00 Length: 98
Ratio: 1.462 Gaps: 3
Percent Similarity: 39.796 Percent Identity: 23.469
```

alignment_block:

US-09-030-606-224/rev x PCT-US94-04496-30 ..

Align seg 1/1 to: PCT-US94-04496-30 from: 1 to: 568

```
250 GGCACATTGTGTCCTCAGTATCAAGACCTTTAAGGCTTC..... 206
||||| : : : : : ||| : : |||||
14 GlyHisArgAlaGlnValArgLysLysProThrValGlnGlyPheThrH1 30
```


205CAT 202
30 SASPTRMETVALPHEVALARGGLYPROGLNHISSEASNIIEGLNHIS 47
201 TY...GAGAAAGTAGATTTCCTTCCTCTCTTTAGACATATGATA 155
47 HEVALGLULYSVALPHEHISLEUHSIGLUSERPHEPROARGPROLYS 63
154 AAATATGCCCTTGTATCCCTTAAACCATGATCCATTTTTTTGATGC 105
64 ARGVALCYS..... 66
104 GGAGTAAGGGGTACACACTCCCTACTACAATGTCACAACTTTTCTTT 55
67LYSASPPTROPTRYLYSVALGLUGLUSERGLTYTYRA 79
54 TGTGTCTCTATTTGGTTGTAAGTACTGTAATTTCTTACACAGAAG 11
79 IAGLYPHEILEUPTROILEGLUVALTYRPHLYSASNLYSGLU 93

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-588-983-16

seq_documentation_block:

; Sequence 16, Application US/08588983
; Patent No. 5854067
; GENERAL INFORMATION:
; APPLICANT: Christopher B. Newgard, et al.
; TITLE OF INVENTION: Methods and Compositions
; TITLE OF INVENTION: for Inhibiting Hexokinase
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,983
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: UTSD:424/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 917 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-588-983-16

alignment_scores:

Quality: 56.50 Length: 67
Ratio: 1.527 Gaps: 5
Percent Similarity: 55.224 Percent Identity: 29.851

alignment_block:

US-09-030-606-224/rev x US-08-588-983-16 ..

Align seg 1/1 to: US-08-588-983-16 from: 1 to: 917

214 AAGGCTTCATTYTGAGAAAGTAGATTTCCTCTCTCT..... 176
551 ARGGLYVALGLUMETHISASNLYSILETYRSERILEPROGLNGLUVALME 567
175TCTCTTTAGACATATGATAAAATATGCCCTT 142
567 THISGLYTHRGLUGLULUPTHEUPHEASPHISILEVALGLN..... 580
141 YTATCCTTAAACCATGTATCCATTTTTTTTG.....ATGGGAGTAA 98
581Cysile,Alaaspheleuglutyrmetcglymetly 592
97 GGGGTACACACTCCCTACTACATGTCACAACTTTTCTTTGTGTGT 49
592 SGLYVALSERLEUPRO.....LEUGLYPHETHRPHESERPHETROCYS 606

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-688-352C-40

seq_documentation_block:

; Sequence 40, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wiegler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 amino acids
; TYPE: AMINO ACID
; MOLECULE TYPE: protein
; US-07-688-352C-40

alignment_scores:

Quality: 55.00 Length: 85
Ratio: 1.222 Gaps: 5
Percent Similarity: 52.941 Percent Identity: 30.588

alignment_block:

US-09-030-606-224/rev x US-07-688-352C-40 ..

Align seg 1/1 to: US-07-688-352C-40 from: 1 to: 384

```
290 TTACCGTGTCCAGATCATGCAACTTTAAAGGAGATGACGCCACATTT 241
    ::||| ||| |||::: ::: |||
305 ValProValProLeuHisSerSerProHisSerProAlaProThr 321
240 GTGTCCTCAGTATCAAGCACCTTTAAGGCTTCATTT..... 201
    ||:::||||| ||| ||| :::
321 rSerSerAlaSerLysThrLeuValLeuLeuSerLeuLeuArgLysL 338
200 .....YGAGAAAGTAGTATTTCTCTCTCTCTCTTTAGACATATGAT 156
    :::::||:::||:::|| ||| |||
338 yslYThrLysSerGlyValPhePheCysPheLeuPhe..... 350
155 AAAATATGCGCTTCTTATCCTTTA...ACCATGTATCCATTTT 109
    ::||| ||| ||| ||| ||| ||| |||
351 .....PheProPheProProAlaProThrHisGlyAlaPhePheTrpAr 365
108 ATGGGAGTAGGGGTACACACTCCCTACTACATGTACAACTTTTC 59
    |||||::: |||::: |||::: |||::: |||
365 gTrpGlyLeu...GlyAsnGly...LeuArgSerArgLysGlyPheT 380
58 TTTT 54
    |||
380 yrphe 381
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-446-875-10

seq_documentation_block:

: Sequence 10, Application US/08446875
: Patent No. 5858751

: GENERAL INFORMATION:

: APPLICANT: Paulson, James C.

: APPLICANT: Wen, Xiaohong

: APPLICANT: Livingston, Brian Duane

: APPLICANT: Gillespie, William

: APPLICANT: Kelm, Soreg

: APPLICANT: Burlingame, Alma L.

: APPLICANT: Medzhradszky, Katalin

: TITLE OF INVENTION: Compositions and Methods for the

: NUMBER OF SEQUENCES: 16

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Poms, Smith, Lande & Rose

: STREET: 2029 Century Park East, Suite 3800

: CITY: Los Angeles

: STATE: CA

: COUNTRY: USA

: ZIP: 90067

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/446,875

: FILING DATE: July 12, 1995

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/102,385

: FILING DATE: August 4, 1993

: ATTORNEY/AGENT INFORMATION:

: NAME: Oldenkamp, David J.

: REGISTRATION NUMBER: 29,421

: REFERENCE/DOCKET NUMBER: 111-197

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (310) 788-5000

: TELEFAX: (310) 277-1297

: INFORMATION FOR SEQ ID NO: 10:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 375

: TYPE: amino acid

: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-446-875-10

alignment_scores:

| Quality: | 54.50 | Length: | 91 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.329 | Gaps: | 5 |
| Percent Similarity: | 45.055 | Percent Identity: | 27.473 |

alignment_block:

US-09-030-606-224/rev x US-08-446-875-10 ..

Align seg 1/1 to: US-08-446-875-10 from: 1 to: 375

```
315 AATGCKAGCTTAAACTGCAACAGTTACCGTGCAGTCATGCAACTY 266
    ||| ::::||| ::: ||| ::: |||
65 AsnLeuAspSerLysLeuProAlaGluLeuAlaThrLysTyraLaAsnph 81
265 TAAAGGAGA.....TGACGGCCA 246
    :: |||
81 eSerGluGlyAlaCysLysProGlyTyraLaSerAlaLeuMetThrAlaI 98
245 CATTGTGCTCCTCAGTATCAAGCACCTTTAAGGCTTCATTTYGAGA 196
    ||| ||||| |||||:::
98 lepheProArgPheSerLysProAlaPromet..... 108
195 AAGTAGTATTTCTCTCTCTCTCTTTAGAACATGATAAATATATGC 146
    |||||::: ||||| |||::: |||
109 .....PheLeu.AsaspSerPheArgLysTrpAlaArgIleArg 122
145 CCTTATTCCTTTAACCATGATCCATTTTGTGATGGGAGTAAG 96
    ||| ||| ||||| |||||:::
122 lupheValPro.....ProPhe.....GlyIleLys 130
95 GTACACACTCCCTACTACAA 75
    ||| ::|||:::
131 GlyGlnAspAsnLeuIleLys 137
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US96-03916-17

seq_documentation_block:

: Sequence 17, Application PC/TUS9603916

: GENERAL INFORMATION:

: APPLICANT: Wild, Martha A.

: APPLICANT: Cochran, Mark D.

: TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS

: NUMBER OF SEQUENCES: 72

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Cooper & Dunham LLP

: STREET: 1185 Avenue of the Americas

: CITY: New York

: STATE: New York

: COUNTRY: U.S.A.

: ZIP: 10036

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US96/03916

: FILING DATE: 23-MAR-1995

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/126,597

: FILING DATE: 24-SEP-1993

: ATTORNEY/AGENT INFORMATION:

: NAME: White, John P.

: REGISTRATION NUMBER: 28,678

: REFERENCE/DOCKET NUMBER: 39116-A

: TELECOMMUNICATION INFORMATION:

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-602-010A-30

seq_documentation_block:

; Sequence 30, Application US/08602010A
; Patent No. 5753235
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,010A
; FILING DATE: February 15, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-602-010A-30

alignment_scores:

| Quality: | 51.00 | Length: | 39 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.889 | Gaps: | 1 |
| Percent Similarity: | 69.231 | Percent Identity: | 28.205 |

alignment_block:

US-09-030-606-224/rev x US-08-602-010A-30 ..

Align seg 1/1 to: US-08-602-010A-30 from: 1 to: 53

```
253 GACGGCCACATTGTGTCCTTCAGTATCAAGCACCTTTAAGGCGTTCCA 204
||| |||::: ||| ::|||:::||||::: |||::: |||
10 AspGlnHisLeuSerValAsnAsnTyrSerGlyThrIleGluPheIleH1 26
203 TTTTGAGAAAGTAGTATTCTCTCTCTCTCTCTTTAGACATATGATAA 154
||||::: |||::: |||::: |||::: |||::: |||::: |||
26 spHeAsnAsnSerCysTyrThrValTyrGlnThrIleGluTyrPhe.... 41
153 AATATGGCCCTTATC 137
||||| |||
42 ..SerCysProArgIle 46
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-680-726A-30

seq_documentation_block:

; Sequence 30, Application US/08680726A
; Patent No. 5804197
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES

; NUMBER OF SEQUENCES: 92

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,726A
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-680-726A-30

alignment_scores:

| Quality: | 51.00 | Length: | 39 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.889 | Gaps: | 1 |
| Percent Similarity: | 69.231 | Percent Identity: | 28.205 |

alignment_block:

US-09-030-606-224/rev x US-08-680-726A-30 ..

Align seg 1/1 to: US-08-680-726A-30 from: 1 to: 53

```
253 GACGGCCACATTGTGTCCTTCAGTATCAAGCACCTTTAAGGCGTTCCA 204
||| |||::: ||| ::|||:::||||::: |||::: |||
10 AspGlnHisLeuSerValAsnAsnTyrSerGlyThrIleGluPheIleH1 26
203 TTTTGAGAAAGTAGTATTCTCTCTCTCTCTCTTTAGACATATGATAA 154
||||::: |||::: |||::: |||::: |||::: |||::: |||
26 spHeAsnAsnSerCysTyrThrValTyrGlnThrIleGluTyrPhe.... 41
153 AATATGGCCCTTATC 137
||||| |||
42 ..SerCysProArgIle 46
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-980-3

seq_documentation_block:

; Sequence 3, Application US/08465980
; Patent No. 5756309
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR470
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; STREET: STUART & OLSTEIN
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey


```

1  COUNTRY:  USA
2  ZIP:  07068
3
4  COMPUTER READABLE FORM:
5  MEDIUM TYPE:  floppy disk
6  COMPUTER:  IBM PC compatible
7  OPERATING SYSTEM:  PC-DOS/MS-DOS
8  SOFTWARE:  PatentIn Release #1.0, Version #1.30
9
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER:  US/08/465,980
12 FILING DATE:  06-JUN-1995
13
14 CLASSIFICATION:  536
15
16 ATTORNEY/AGENT INFORMATION:
17 NAME:  Ferraro, Gregory D.
18 REGISTRATION NUMBER:  36,134
19 REFERENCE/DOCKET NUMBER:  325800-446
20
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE:  201-994-1700
23 TELEFAX:  201-994-1744
24
25 INFORMATION FOR SEQ ID NO:  3:
26 SEQUENCE CHARACTERISTICS:
27     LENGTH:  247 amino acids
28     TYPE:  amino acid
29     STRANDEDNESS:
30     TOPOLOGY:  linear
31     MOLECULE TYPE:  protein
32
33 US-08-465-980-3

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 52.50 | length: 77 |
| Ratio: | 1.280 | Gaps: 4 |
| Percent Similarity: | 53.247 | Percent Identity: 25.974 |

alignment_block:
US-09-030-606-224/rev x US-08-465-980-3 .

Align seg 1/1 to: US-08-465-980-3 from: 1 to: 247

```

259 AGGATGGACCGGCACATTGTGTCTCCTTCAGTAT.....CAAAG      222
    |||::|||::|||::|||::   ::   ::|||
39 ArgLeuaspsSerHisLseUHistHrPrometYrrLeupeUeSeraNle     55
    |||::|||::|||::|||::   ::   ::|||
221 CACCCTTAAGGGCTCCCATTTGYGAGAAGTAGTATTCTCTCCTCTC       172
    :::|||::   ::   |||:::|||   :|||::|||::||
55 userPheSerAspleucysPheserSeVal...ThriLeProlysl        70
    |||::|||::|||::|||::   ::   |||:::|||
171 TTTTAGAACATATGATAAATATATGCCCTTYTATCCTTAACCATGTATC     122
    |||::|||::|||::|||::   ::   |||:::|||
70 euleUGlnasmeGlinsnglaspProserile.....            81
    |||::|||::|||::|||::   ::   |||:::|||
121 CATTTTTTTTTTGATGGGAGTAGAGGGGTACACACTCCCTACTACAATGT     72
    |||::|||::|||::|||::   ::   |||:::|||
82 .....ProtYrAlaaSpCy          86
    |||::|||::|||::|||::   ::   |||:::|||
71 CACAAC.....TTTTCTTTTTGTGTC         47
    ::   ::   |||::|||::|||::|||
86 sleuthrglmetyrrphephelouleuphe           96
```


This Page Blank (uspto)

JOURNAL Science 273 (5278), 1058-1073 (1996)
MEDLINE 96337999
REFERENCE 2 (bases 1 to 10094)
AUTHORS Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,
Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
Merrick,J.M., Glodek,A., Scott,J.D., Geoghegan,N.S., Weidman,J.F.,
Fuhmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M.,
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
Smith,H.O., Woese,C.R. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1996) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Jan 30, 1998 this sequence version replaced gi:1591965.
FEATURES
source location/Qualifiers
1. 10094
/organism="Methanococcus jannaschii"
/db_xref="taxon:2190"
95. 364
/gene="MJ1325"
95. 364
/gene="MJ1325"
/note="similar to GB:L10909 SP:P37374 PID:152977 percent
identity: 39.74; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cadmium resistance protein"
/protein_id="AAB99335.1"
/db_xref="GI:1592329"
/db_xref="PID:g1592329"
/translation="MEKYEKAIEFKAFGDPTRLMLKLLAENGSMCVCKITDELKPP
OPTISHLNLILKAGIVKARKEGTWTFYIVDDRKEIILVDEL"
564. 1739
/gene="MJ1326"
564. 1739
/gene="MJ1326"
/note="similar to SP:P17103 PID:43450 percent identity:
50.68; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein/ member of GTP1/OBG-family"
/protein_id="AAB99336.1"
/db_xref="PID:g1591967"
/db_xref="GI:1591967"
/translation="MLFLSLKRIFFRNHKNFVGVGIEEIRIEELKKTYPN
KATQKHIGRLAKIAKLREQAQSGGGGKGAVAKSGDATAFVGFPVSGSTLLN
KLTNAKSEVGAYAFETTLITVPGILEYKGAQIQLDAPGITVGASSGKRGTEVLSAVR
SADLILIVDIYTLIDHLPVLEKELYNVGIQLDQTPPDVAKIVKERGIVSSSTPELTH
IDEDTIEALINEYRINADVIRNEDITLQFIDVAGNRVYIPSLVVNKKIDLADDEY
LKIYKQLEEFEGKDYILVSGNGINLIDLEKIEKLGFIKIYLPQGGKPPDEPLI
MRGATVVDYCEKHLKDFVRNFRYAQVWGSAKHPPGQRVGLDHKLEDGDLITVIKR"
1748. 2485
/gene="MJ1327"
1748. 2485
/gene="MJ1327"
/note="invalid gene; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1327"
/protein_id="AAB99343.1"
/db_xref="PID:g1591968"
/db_xref="GI:1591968"
/translation="MVVKLVNNELKILSGKLDIFREIYNKIKNDENINDRNLDDKYL
SLKDYTISEENLKIKFSEPKDEISYEGRTPYDLTCYGINGKNFLIFINNKFGDL
KSNTRNDVTTYNNLLRLYLGIKQRLSEITINGELVYNRISGENEIVSYGLFVVDKYR
RGYKFFLEELIKDDFYVNPNNMFQIRYSPNLGDPIDYFAVKKLIDAILSELSKSLN
SIKTEILVNSIKIQLINIKEKGHGED"
3148. 4044
/gene="MJ1328"
3148. 4044
CDS

/gene="MJ1328"
/note="similar to GB:L42023 SP:P44414 PID:1003910
PID:1222446 PID:1204764 percent identity: 30.63;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="modification methylase, type II R/M system"
/protein_id="AAB99337.1"
/db_xref="PID:g1591969"
/db_xref="GI:1591969"
/translation="MQKNQITFLNTWKFLTLHQKPMSTFTFKIPDEPHVLKDIKAV
GVGLVSGFDEAFLLNEDDISKLNDEKQLKNEFKAKNCKRFVEGEFVQYILLIEDNLK
DEEITFKTKYPNIYKKLKEKDRMENRYLPKRNKWFENQALRNYKFLNKKRIYVP
TIDRKPNRFSLGDELLPSGDVIFIQPYNDDIYFLLGLYLNSSFFRNYLANGGRG
GRVAFQKLENAKIPFESDEVKEKIKNIYDIIYNLKNKDIENLERQIDYIIVSAI
NNNQFKGYQTTLKNLKPRLKG"
4049. 4933
/gene="MJ1329"
4049. 4933
/gene="MJ1329"
/note="similar to GB:M26978 SP:P22624 PID:551896 percent
identity: 50.90; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="methionyl aminopeptidase (map)"
/protein_id="AAB99338.1"
/db_xref="PID:g1591970"
/db_xref="GI:1591970"
/translation="MEIEGYEKIIEAGKIAKRVREAVKLIXPGVKLLEVAEFVENRI
RELGEPAFPCNISINEIAAHYTPKLNDELFEKDDVVKLIDGAHVGDIADTAITVD
LSNSYKDLVKA SEDALYTVIKEINPPMNTGEMGKIIQEVIESGYRPISNLSGHVMHR
YELHGISIPNVERENQYIDVGLVAIEPFTDGEFVMDKNGNLYKFLAKRPIRL
PQARKLIDVISKNYPYLPFAERWVLKNESERLALNSLIRASCIYGPILKERENGIVG
QAEHTIILITENGVEITTK"
5079. 5462
/gene="MJ1330"
5079. 5462
/gene="MJ1330"
/note="similar to GB:AE000666 percent identity: 54.47;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB99339.1"
/db_xref="PID:g2826396"
/db_xref="GI:2826396"
/translation="MDAGMDLYIGSNEERDRYKEKLEIKQLSNPNVSTLLIAIT
LDNGGRANLPLFNYNEDPNYVYVDEVIGLAIAINEIAGTKAIFNFRYDAKPGIIGEL
DRKFMFLDDAIGLAGCMSKVFEE"
complement(5465. 5899)
/gene="MJ1331"
complement(5465. 5899)
/gene="MJ1331"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1331"
/protein_id="AAB99344.1"
/db_xref="PID:g1591972"
/db_xref="GI:1591972"
/translation="MVASMATYELRIYGNVCAEFIDKVESTLGKLDVNGVVYVKD
SVRIANFPNEKKRQLFKELIKDLEDGGLIKVERIEERDLNTYIEFPNGLNKISTNE
LKEINKLIDKTIISYLENIFENALEKQIKVSEIRDILKDTFEV"
complement(5908. 7089)
/gene="MJ1332"
complement(5908. 7089)
/gene="MJ1332"
/note="similar to SP:P42942 PID:1165215 PID:1323377
PID:790501 percent identity: 39.69; identified by sequence
similarity; putative"
/codon_start=1
CDS

CDS

/transl_table=11
/product="GTP-binding protein, member of GTP1/OBG-family"
/protein_id="AAB99340.1"
/db_xref="PID:g1591973"
/db_xref="GI:1591973"
/translation="MAMIGLVKPNVGKSTMENALTEPKAEIGNYPFTTIQDNKGIAIYITSPCKELGVKCNPNSNKCIDGRIPVEVIDVAGLPGAHEGRGMGNRFLDLROADAFILVVDA SGKTDAEGNPTENDPVEDVKFLLNELDMWISILTNNWDKLARRAQOEKNIVKALKDQLSGLNIDEDDIKAIRDMDESPIKWTEEDLI AKLRISKPMIIAAANKADHPDAEKNIERLKKEFKDYIVIPTS AEIELAKRAEKAGIIKRKENDEFIIDESKVNEQMRRAPFDYIKDFLKKYGGSVQCIN KAYFDLDIMIVVPVEDENKFSDKQGNVL PDAFLVKKGSTARDLAYKRVHTELGEKFIYA IDAKKKIRVGADYE LKHNDLIKIVSAAK"

gene

7182..7436
/gene="MJ1333"
7182..7436
/gene="MJ1333"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1333"
/protein_id="AAB99345.1"
/db_xref="PID:g1591974"
/db_xref="GI:1591974"
/translation="MPSPKIQEITNELDNLMMNRERKYIELVATVEYL NLTLEPSKREKFKREALVAETVEDVELIKAIRKLGMGARRYLLTLEG"

gene

complement(7495..7908)
/gene="MJ1333.1"
complement(7495..7908)
/gene="MJ1333.1"
/note="Brute Force ORF; identified by Genemark; putative"

| | | | | |
|---------------------------|--------|-----------------|-----------|---------------|
| Query Match | 14.18; | Score 45; | DB 2; | Length 10094; |
| Best Local Similarity | 48.28; | Pred. No. 0.38; | | |
| Matches 120; Conservative | 2; | Mismatches 127; | Indels 0; | Gaps 0; |

| | | | | | | | | | | | | |
|----|------|------------|---------|-----------|-----------|-----------|------------|------------|-----------|-----------|---------|------|
| QY | 5 | TGAAGGCTTC | TGTTAG | AAAAATAG | TACAGTTAC | AACCAATAG | GAACAAC | AAAAA | GA | AAAAA | 64 | |
| | | | | | | | | | | | | |
| Db | 3438 | TGAAGACAAC | CTAA | AAGATGA | AGAAATAT | TTAA | AAACCAATAT | CCAAACATAT | ATA | AAAAA | 3497 | |
| QY | 65 | GTTTGTGAC | ATTGTAG | TAGGAGTGT | TACCCCTT | ACTCCCAT | CAAAAAA | AAAAATG | GAT | 124 | | |
| | | | | | | | | | | | | |
| Db | 3498 | ATTGTTGA | AGTTAA | AGACAGA | ATGAGA | ATAGATAC | CTTCC | CAAAAAA | CAAAAAATG | GTT | 3557 | |
| QY | 125 | ACATGTTAA | AGATARA | AGGCCA | TATTTT | TATCATAT | GTCTAA | AAAGAGA | GAGAGA | 184 | | |
| | | | | : | | | | | | | | |
| Db | 3558 | TAATTTGG | CAGCAT | TGAGAA | CTATAA | ATTTT | TATATA | AAAAA | ATCTAA | ATAAAAACG | AT | 3617 |
| QY | 185 | AAATACTAC | TTTCTC | RAAATG | GAAGCC | CTTAA | AGGTGC | TTGAT | ACTG | AAAGACACA | AT | 244 |
| | | | | : | | | | | | | | |
| Db | 3618 | ATATGTT | CCAAC | TTTAG | ACAGAA | AAACCAT | ATATA | TAGGTT | AGTTAG | GAGATG | ATGAACT | 3677 |
| QY | 245 | GTGCGCG | TGC | 253 | | | | | | | | |
| | | | | | | | | | | | | |
| Db | 3678 | ATTACCT | TC | 3686 | | | | | | | | |

| RESULT | 2 |
|------------|---|
| AC006755 | |
| LOCUS | |
| DEFINITION | AC006755 199917 bp DNA HTG 23-FEB-1999 |
| | Caenorhabditis elegans clone Y40C5, WORKING DRAFT SEQUENCE, 1 |
| | unordered pieces. |
| ACCESSION | AC006755 |
| NID | g4263183 |
| VERSION | AC006755.1 GI:4263183 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | Caenorhabditis elegans. |
| ORGANISM | Caenorhabditis elegans |
| | Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; |
| | Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis. |
| REFERENCE | 1 (bases 1 to 199917) |
| AUTHORS | Waterston, R.H. |

```

TITLE      The sequence of Caenorhabditis elegans clone
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 199917)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (23-FEB-1999) Genome Sequencing Center, Washington
                                University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                                MO 63108, USA
COMMENT    * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 199917: contig of 199917 bp in length.
FEATURES   Location/Qualifiers
            1..199917
               /organism="Caenorhabditis elegans"
               /db_xref="taxon:6239"
               /clone="Y40C5"
BASE COUNT 63887 a 33980 c 34366 g 65836 t 1848 others
ORIGIN

```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|----------------|
| Query Match | 13.7%; | Score 43.8; | DB 34; | Length 199917; |
| Best Local Similarity | 51.7%; | Pred. No. 0.52; | | |
| Matches 90; | Conservative 2; | Mismatches 82; | Indels 0; | Gaps 0; |

QY 30 ACAGTTACAAACCAATAGCAACAACAAAAAGAAAGTTGTGACATTGTAGTAGGAGCTG 89
| | | | | | | | | | | | | | | | | | | | | |
Db 127380 AAAATTGAAACTATTTGTATGAAAAAATGGCAAAACCTTGGCAGGTTGGCCAAGAGTTG 127439
QY 90 TGTACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGATARAAGGGCAA 149
| | | | | | | | | | | | | | | | | | | | | |
Db 127440 TGTAAAGCTTTTTCACGTTTAAACGAAGTTTATATACCTTGAATAATGAACAAATTA 127499
QY 150 TATTTATCATATGTTCTTAAAGAGAAGAGAAGAAAATACTACTTTCRRAA 203
| | | | | | | | | | | | | | | | | | | | | |
Db 127500 GTTTATGCCAANTACTCTGTAAAAAATNTACTTAANCAATTTATACCTTGGATMAAA 127553

| | | | | | |
|------------|---|---------------|-------------------------|--------|-------------|
| RESULT | 3 | | | | |
| AE001391/c | | | | | |
| LOCUS | | | | | |
| DEFINITION | AE001391 | 14553 bp | DNA | INV | 06-NOV-1998 |
| | Plasmodium falciparum | chromosome 2, | section 28 of 73 of the | | |
| | complete sequence. | | | | |
| ACCESSION | AE001391 | AE001362 | | | |
| NID | 93845168 | | | | |
| VERSION | AE001391.1 | GI:3845168 | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| | malaria parasite P. falciparum. | | | | |
| | Plasmodium falciparum | | | | |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | | | |
| AUTHORS | 1 (bases 1 to 14553) | | | | |
| | Gardner,M.J., Tettein,H., Carucci,D.J., Cummings,L.M., Aravind,L., | | | | |
| | Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., | | | | |
| | Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., | | | | |
| | Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., | | | | |
| | Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L. | | | | |
| TITLE | Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum | | | | |
| JOURNAL | Science | 282 (5391), | 1126-1132 | (1998) | |
| MEDLINE | 99021743 | | | | |
| REMARK | Erratum: [[published erratum appears in Science 1998 Dec 4;282(5395):1827]] | | | | |
| REFERENCE | 2 (bases 1 to 14553) | | | | |
| AUTHORS | Gardner,M.J. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA | | | | |
| FEATURES | location/Qualifiers | | | | |

[illegible]

| | | | | | |
|-----------------------|---|---|--------------------------------------|----------------|--|
| BASE COUNT | 5522 a | 1421 c | 1141 g | 6469 t | Y NVLTKNGNNNSPSYIFLNSNRNNNNFLRRKENYQVLLLYNPKFNDGAKRNSYILL ASHSKNRKRNKINHNINITKREVSRDRDDEDDNYDDDDDDDDDDDDDDDDDDDDDD DENTYDDDDDDDDGENTYEDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD PYETOKSKIRSKDENVAHONGGNTLSLENYKIKKANSDLTDKKNKVKKODEMD ETNKNKNETNEKDEKDEKEEDNDINDDIMEDEYEDYITEEHKIDHKICSEK MKNVYEFLEKESYRFNLNNVSNDFEENEKVINERYTYVKHICHIKKENLFTIPYD PYVNFLYOHFIKEFDELKIFYKDKSLYAIIPPISENLKNEIKMKIKRKIEDSKVTLR TVRKQMDKLEKFKNKIGKDIYFKOKNYIOSIHDQTKNIEKLFADTK" |
| ORIGIN | | | | | |
| Query Match | 13.5%; | Score 43.2; | DB 37; | Length 14553; | |
| Best Local Similarity | 57.7%; | Pred. No. 0.91; | | | |
| Matches | 75; | Conservative | 1; | Mismatches 54; | Indels 0; |
| | | | | Gaps | 0; |
| QY | 38 | AACCAATAGGACACACAAAAAGAAAGTTGTGCACATTGTAGTAGGAGTGTACCCC | 97 | | |
| Db | 6985 | AACCAACAAAAAATAAATAACAAAGTTGAAGAATTATTGGAAGAAAGTTAAATT | 6926 | | |
| QY | 98 | TTACTCCCATCAAAAAAATGATACATGTTAAAGGATARAAGGCAATATTAT | 157 | | |
| Db | 6925 | GTAAGAAGGATAAAAAATAATATATATATAATAAGAAAAAGCTAAATTTTAC | 6866 | | |
| QY | 158 | CATATGTCT | 167 | | |
| Db | 6865 | TTTTTTTTTT | 6856 | | |
| RESULT | 4 | | | | |
| AC005140/c | | | | | |
| LOCUS | AC005140 | 194410 bp | DNA | HTG | 01-APR-1999 |
| DEFINITION | Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 14 | | | | |
| ACCESSION | AC005140 | | | | |
| NID | g4558572 | | | | |
| VERSION | AC005140.3 | GI:4558572 | | | |
| KEYWORDS | HTG; HTGS_PHASE1. | | | | |
| SOURCE | malaria parasite P. falciparum. | | | | |
| ORGANISM | Plasmodium falciparum | | | | |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | | | |
| AUTHORS | 1 (bases 1 to 194410) | | | | |
| TITLE | Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W. | | | | |
| REFERENCE | Plasmodium falciparum 3D7 chromosome 12 | | | | |
| JOURNAL | Unpublished | | | | |
| AUTHORS | 2 (bases 1 to 194410) | | | | |
| TITLE | Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W. | | | | |
| JOURNAL | Direct Submission | | | | |
| COMMENT | Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | | | | |
| | On Apr 2, 1999 this sequence version replaced gi:4337162. | | | | |
| | * NOTE: This is a 'working draft' sequence. It currently | | | | |
| | * consists of 14 contigs. The true order of the pieces | | | | |
| | * is not known and their order in this sequence record is | | | | |
| | * arbitrary. Gaps between the contigs are represented as | | | | |
| | * runs of N, but the exact sizes of the gaps are unknown. | | | | |
| | * This record will be updated with the finished sequence | | | | |
| | * as soon as it is available and the accession number will | | | | |
| | * be preserved. | | | | |
| | 1 | 13954: contig of 13954 bp in length | | | |
| | * | 13955 | 14154: gap of unknown length | | |
| | * | 14155 | 32431: contig of 18277 bp in length | | |
| | * | 32432 | 32631: gap of unknown length | | |
| | * | 32632 | 105078: contig of 72447 bp in length | | |
| | * | 105079 | 105278: gap of unknown length | | |
| | * | 105279 | 117575: contig of 12297 bp in length | | |
| | * | 117576 | 117775: gap of unknown length | | |
| | * | 117776 | 120340: contig of 2565 bp in length | | |
| | * | 120341 | 120540: gap of unknown length | | |
| | * | 120541 | 132816: contig of 12276 bp in length | | |
| | * | 132817 | 133016: gap of unknown length | | |
| | * | 133017 | 135389: contig of 2373 bp in length | | |

| FEATURES | source |
|-----------------------------------|---|
| * 135390 | 135589: gap of unknown length |
| * 135590 | 165594: contig of 30005 bp in length |
| * 165595 | 165794: gap of unknown length |
| * 165795 | 169039: contig of 3245 bp in length |
| * 169040 | 169239: gap of unknown length |
| * 169240 | 172055: contig of 2816 bp in length |
| * 172056 | 172255: gap of unknown length |
| * 172256 | 173886: contig of 1631 bp in length |
| * 173887 | 174086: gap of unknown length |
| * 174087 | 179217: contig of 5131 bp in length |
| * 179218 | 179417: gap of unknown length |
| * 179418 | 186342: contig of 6925 bp in length |
| * 186343 | 186542: gap of unknown length |
| * 186543 | 194410: contig of 7868 bp in length. |
| location/Qualifiers | |
| 1. .194410 | |
| /organism="Plasmodium falciparum" | |
| /db_xref="taxon:5833" | |
| /chromosome="12" | |
| BASE COUNT | 77243 a 19864 c 17424 g 77256 t 2623 others |
| ORIGIN | |

| Query Match | 12.7% | Score 40.8; | DB 35; | length 194410; |
|-----------------------|--|-----------------|-----------|----------------|
| Best Local Similarity | 47.2%; | Pred. No. 2.4; | | |
| Matches 117; | Conservative 2; | Mismatches 129; | Indels 0; | Gaps 0; |
| QY 20 | AGAAATAGTACAGTTACAACCAATAGGAACAACAAAAAGAAAAAGTTGTGACATTGTA | 79 | | |
| Db 40205 | ATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAA | 40146 | | |
| QY 80 | GTAGGGAGTGTGTAACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGGAT | 139 | | |
| Db 40145 | AAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAAGAT | 40086 | | |
| QY 140 | ARAAGGCAATATTTTATCATATGTTCTTAAAAAGAGAAGAGAAATACTACTTCTC | 199 | | |
| Db 40085 | GAATGATTAATTTTATGATAATTTATTTGTTTATAAAAAATAATATACATATATATA | 40026 | | |
| QY 200 | RAAATGGAAGCCCTTAAAGTGCTTTGATTAAGAGACACAATGTCGCCGTCATCCT | 259 | | |
| Db 40025 | TATGTTAAAT | 39966 | | |
| QY 260 | CCTTTARA 267 | | | |
| Db 39965 | TCCTTAAA 39958 | | | |

| | |
|------------|--|
| RESULT | 5 |
| AC006279/c | |
| LOCUS | |
| DEFINITION | AC006279 219200 bp DNA HTG 01-APR-1999 Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 6 unordered pieces. |
| ACCESSION | AC006279 |
| NID | g4558576 |
| VERSION | AC006279.3 GI:4558576 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | malaria parasite P. falciparum. Plasmodium falciparum |
| ORGANISM | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. |
| REFERENCE | 1 (bases 1 to 219200) Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W. |
| AUTHORS | Plasmodium falciparum 3D7 chromosome 12 |
| TITLE | Unpublished |
| JOURNAL | 2 (bases 1 to 219200) |
| REFERENCE | Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology |
| JOURNAL | Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA |
| COMMENT | On Apr 2, 1999 this sequence version replaced gi:4337165. * NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces |

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
  1
  77527 77526: contig of 77526 bp in length
  77727 77726: gap of unknown length
  140132 140131: contig of 62405 bp in length
  140332 140331: gap of unknown length
  173509 173508: contig of 33177 bp in length
  173709 173708: gap of unknown length
  212797 212797: contig of 39089 bp in length
  212998 212997: gap of unknown length
  214905 214904: contig of 1907 bp in length
  215105 215104: gap of unknown length
  219200 219200: contig of 4096 bp in length.
  location/Qualifiers
  1. 219200
  /organism="Plasmodium falciparum"
  /db_xref="taxon:5833"
  /chromosome="12"
BASE COUNT 88210 a 22418 c 19532 g 88036 t 1004 others
ORIGIN

```

| Query Match | 12.7% | Score 40.8; | DB 35; | Length 219200; |
|-----------------------|--|-----------------|-----------|----------------|
| Best Local Similarity | 47.2%; | Pred. No. 2.4; | | |
| Matches 117; | Conservative 2; | Mismatches 129; | Indels 0; | Gaps 0; |
| QY 20 | AGAAATAGTACAGTTACACCAATAGGACACACAAAAAGTTGTGACATTGTA | 79 | | |
| Db 85299 | ATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAA | 85240 | | |
| QY 80 | GTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGGATACATGTTAAAGGAT | 139 | | |
| Db 85239 | AAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAAGAT | 85180 | | |
| QY 140 | ARAAGGCAATATTTTATCATATGTCTTAAAGAGAGAGAAATACTACTTCTC | 199 | | |
| Db 85179 | GAATGATTAAATTTTATGATAATTATTGTGTATTAATAATAATATACATATATATA | 85120 | | |
| QY 200 | RAATGGAAGCCCTTAAAGGTGCTTGATACCTGAAGACACAATGTGGCCGTCCACT | 259 | | |
| Db 85119 | TAT | 85060 | | |
| QY 260 | CCTTTARA 267 | | | |
| Db 85059 | TCCTTAAA 85052 | | | |

| RESULT | 6 |
|------------|--|
| LOCUS | SADNAMUPR/c |
| DEFINITION | SADNAMUPR 4013 bp DNA |
| ACCESSION | S.aureus plasmid encoded DNA, mup R gene. |
| NID | X75439 |
| VERSION | g438226 |
| KEYWORDS | X75439.1 GI:438226 |
| SOURCE | Isoleucyl-tRNA synthetase. |
| ORGANISM | Staphylococcus aureus. |
| REFERENCE | Staphylococcus aureus |
| AUTHORS | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; |
| TITLE | Staphylococcus. |
| JOURNAL | 1 (bases 1 to 4013) |
| MEDLINE | Hodgson,J.E., Curnock,S.P., Dyke,K.G., Morris,R., Sylvester,D.R. |
| REFERENCE | and Gross,M.S. |
| AUTHORS | Molecular characterization of the gene encoding high-level |
| TITLE | mu1pirocin resistance in Staphylococcus aureus J2870 |
| | Antimicrob. Agents Chemother. 38 (5), 1205-1208 (1994) |
| | 94346838 |
| | 2 (bases 1 to 4013) |
| | Hodgson,J.E. |
| | Direct Submission |

JOURNAL Submitted (12-OCT-1993) J.E. Hodgson, Smithkline Beecham Pharmaceuticals, Brockham Park, Betchworth, Surrey, RH3 7AJ, UK

FEATURES

source location/Qualifiers 1..4013

CDS

/organism="Staphylococcus aureus"
/strain="J2870"
/db_xref="taxon:1280"
/clone="POX 301"
<1..450
/note="ORF B"
/codon_start=1
/transl_table=1
/protein_id="CA53190.1"
/db_xref="PID:g438227"
/db_xref="GI:438227"
/db_xref="SWISS-PROT:P41369"
/translation="EPEENHVDLILNLTSGYKKNDLKNLRLNSYTSLKGGIILID
LMNPAFFYKNFEQMIENFKDYENQTIKINKIYNSIVSTWDYKKNNEYSKDLSRELK
FKLYTTNDIICSSLYNFTLSNISFYGDMKGNFHENSPRLIVKITK"
410..3588
/gene="1les"
410..415
/gene="1les"
434..439
/gene="1les"
461..466
/gene="1les"
/bound_mofety="promoter"
477..3551
/gene="1les"
/codon_start=1
/transl_table=1
/product="isoleucyl tRNA synthetase"
/protein_id="CA53189.1"
/db_xref="PID:g581558"
/db_xref="GI:581558"
/db_xref="SWISS-PROT:P41368"
/translation="MTKKYLNTQNEISAFWNTQKIFKKSIDNRKGOESFVYDGPPTA
NGLPHAGHVLGRVTKDLVARLKTMOGFYERKAGWDTGLPVELEYEKIGIKGKODI
EKYGIENFINECKSVFNEYEKERDFSKDIGVWDMSPYITLNNYIESVWNILSTF
HKGLLYKGHKVPYCTHDOTALSHEVAQGYKNVCDLSAVKPOLNSKDTYFLSWT
TTPWTLPANVALAINKDLNYSKIRVENEYIILATDLINSIITEKEIIDTFSSNLIN
LKYIPFESDGLVNAVYVVDGEFVTNSEGIVHIAIPAHGEDDYQLVLERDLFLNVI
TREGVYNDRFPELVGNKAKNSDIEIILKLSKQLLYKKOKYEHNYPHCWRCGNPLIY
AMEGWFIKTTNEKNEIINNNNNIEMFSPSHKEGRMGNFLENVDMNIGRNRWGTPLN
VWICNDCNHEYAPSSIKDLQNNNSINKIDEDIELHRPYVDNITLCPKCNCKMSRVEEV
IDVWFDGSGMPFAQHYPFDNOKIFNQHPADFAEGVDQTRGFYSLVISTILKKG
SSYKRALSLGHILDSNGKMSKSGNVINPTLINSYGADSLRNLISDSAPWNNKRE
SENIVAQTSKSFIDTLNITYKEYNMNKKIDHYNPNNEITKSRLTDNVALSRLNTLIK
ESNIYVNNYDFTSARLINEYNTNISNMYIGDSRGFRWEGISNDKDAYNTLYEILT
TLSRLVAPFVPEIIEKIHYNLTGKSVHLQDYPQYKESFINQALEDEMHVIVIKIVELSR
QARKNADLKIKOPLSKMVIKPNSQLNSFLPNYSIIDEKDELNIKNIETDNDINDYIT
ELKLNFSVGPPLGKNTKNIQTLDLSLEYDKSLIESNNFKSSDAELTKDDFIIR
TLPKDSYQLSIEDNDCVILLDKNLSPELIREGHARELIRLIQQLAKKKNLPIINQRIDY
IGVTGELLESIKITNKNMKEKNFVIKNIHNVIDEYENTIHFNKEIKISILY"

CDS

stem_loop
3557..3588
/gene="1les"
3621..>4013
/note="ORF C"
/codon_start=1
/transl_table=1
/protein_id="CA53191.1"
/db_xref="PID:g438228"
/db_xref="GI:438228"
/db_xref="SWISS-PROT:P41370"
/translation="MYNEDYSKLPKIKNIQKIFPIAGGYVNLSESYDASNKYFLKLOP
NTKSNFEDYELSLKEITDKNIPVQIINKGELDNNSFLLEFIENGHAYPESYRKLK
KIYAMHKNINSLNLEGFSHNFGTIEF"

BASE COUNT 1591 a 506 c 618 g 1298 t
ORIGIN

Query Match

12.6% Score 40.4; DB 1; Length 4013;

| | | | |
|-----------------------|------|--|----------------------------------|
| Best Local Similarity | | 52.4%; | Pred. No. 4.3; |
| Matches | | 86; Conservative 1; | Mismatches 77; Indels 0; Gaps 0; |
| OY | 13 | TCCTGTAGAAAATAGTACAGTTACACCAATAGGACACACAAAAAGTTGTGA | 72 |
| Db | 3671 | TATTTTTCATATATTTTAAATGCTGTAATTTAGATAATCAAAATATACATAGTTATCA | 3612 |
| OY | 73 | CATGTAGTAGGAGTGTGACCCCTTACTCCCATCAAAAAAATGATACATGTT | 132 |
| Db | 3611 | CCCTTGCTTGCTTGAATACATCATGCGCACCTACTAAATAGAGTGGCATGTTGT | 3552 |
| OY | 133 | AAAGCATARAAGGCCAATATTTATCATATGTCTAAAGAGAA | 176 |
| Db | 3551 | TTAATATAATAAGGAAATTTTATTCTTTATTTAAATGAA | 3508 |

RESULT 7

CEFA6C3 29777 bp DNA INV 23-NOV-1998
LOCUS
DEFINITION Caenorhabditis elegans cosmid F46C3, complete sequence.
ACCESSION Z66563
NID g1051319
VERSION Z66563.1 GI:1051319
KEYWORDS HTG; Myosin; Protein kinase.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Wilson,R., Alnscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E., Staden,K., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohlman,P.

TITLE
JOURNAL
MEDLINE
COMMENT
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
Nature 368 (6466), 32-38 (1994)
94150718
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cg1-bin/display?db=wormace&class=Sequence&object=F46C3

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F46C3. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone F46C3 is at 1 in this sequence. The true right end of clone F46C3 is at 9379 in sequence Z68107.
The true left end of clone F46C5 is at 29718 in this sequence. The

true right end of clone F08B12 is at 11562 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268104.
The end of this sequence (29718..29777) overlaps with the start of sequence 268107.

FEATURES

source

location/Qualifiers
1..29777
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F46C3"
complement(5728..7331)
/gene="F46C3.2"
complement(join(5728..5791,5836..6003,6061..6189,6239..6419,6467..6608,7046..7192,7242..7331))
/gene="F46C3.2"
/codon_start=1
/protein_id="CAA91468.1"
/db_xref="PID:e1346869"
/db_xref="PID:g3877185"
/db_xref="GI:3877185"
/db_xref="SPTREMBL:Q20457"

gene

CDS

/translation="MWLVVGFQLPVKLTLLKGGMLRMQLTTHAVDSVILDYVTGM
DKHQRALNLRREGVLTDECFINISTHVGPEROVMPRVVEAMRIIOEGEODLOKE
KPMCLKIFTMKHYSSASFDNHYRADLAKVSLTAARYTMPITWYLFESFMTLATFASHI
WSQVTNYSLOETLSMTLPTIILVSSSLIGMLLYFNVMGFLCCENNESMRLVTK
DLDYSSLYVHFCENKFDLRIFYVKGLQFILMTYMLLAVYAHEDLNITYPALVCVLT
NGFGLMFARMAEFFRRRCRTRDGTSIAR"
complement(9727..26596)
/gene="hum-4"
complement(join(9727..9807,10085..10186,10257..10423,10474..10616,11034..11173,11223..11383,11430..11604,11647..11780,11846..11979,12137..12416,12469..12577,12628..12776,12901..13082,13137..13261,13861..13979,14025..14261,14311..14431,14665..14838,14887..15013,15220..15302,15960..16008,16080..16201,16482..16676,16726..17201,17383..17609,17720..17873,17922..18046,18094..18288,18343..18477,18522..18682,18876..19118,19162..19930,19990..20160,20230..20339,20385..20525,20573..20737,20784..21543,21589..21783,21827..21993,23751..24069,25817..25911,26033..26348,26397..26596))
/gene="hum-4"
/note="Similarity to myosin; cDNA EST yk249a4.5 comes from this gene; cDNA EST yk270h6.5 comes from this gene; cDNA EST yk284c3.5 comes from this gene; cDNA EST yk353c4.5 comes from this gene; cDNA EST yk470b4.5 comes from this gene; cDNA EST yk249a4.3 comes from this gene; cDNA EST yk470b4.3 comes from this gene"

gene

CDS

/codon_start=1
/protein_id="CAA91469.1"
/db_xref="PID:e1346870"
/db_xref="PID:g3877186"
/db_xref="GI:3877186"
/db_xref="SPTREMBL:Q20456"
/translation="MSTEVIALINPDHTEFDHIEANISRKNSIACYNMOTIPLNF
NFSKPAIFKISVAFGDLRLPINYDPEATHAPRTNMTATNPYSTSPRFSRGEIA
VNNLNECSLKMROHTDHOIFNKNLVTKCEDVQRLGNSHNLQSFEPDYLLEGIVAN
IHFLSEEFYPSSEETIFLSETAEKFKKHVPKLFRCRSSPEIYSINRMCIMARWVN
EGYVMSDNFTRLMSDTRWEVETVNOFILRWYRDRGRLDLERFELLQSDSTRIGVSE
EMESSTPIHGFYKERTSKPSWLASLILKPKQPTKASNVLVFENDDKDSEQLSW
ETSSSTGVNAVAKNALNKIFNMSNAESIYFGESGSGSKSYNFAFKYLTQSQSKY
STKHSSSIEFVEKSGCAKTLKNDDEATREGSIDLLYKRNVLTGLNLKYTPLEVPYV
ISOKPERNFENIFYEYHGLSDEMAKFGIKGLQKFYINOGNSSENIOHDVNRFKHL
ESALHVLGFSDHCHMSYKIIISTILHIGNIYFTRNPNVEDVVEIKLAEIKWIAF
LLEVDFOVLKFLPTSEDGSTIELNALDNDSFAMMIYEELFKVNLNRIKGLKCS
LHTGVISILDHYGFEKYNNGNVEEFLINSVNERIENLFVKHCFHODULIDYAKGISVD
YKVPNSIENGKTVELLEKKPYGLSLTLTDECKPKGTHETYLEHCNLNHTDRSAYGA
RNKERLEFGVRHCIGTTWNTDEFARNRKRLISLSAVOLMRNSKNPIIGLLFESYGN
TSDIIVSOAQFVLKGAQDIADKINVSHVHVCIKSNNEROSTKFDIPLVNQIKNLL
LAELLSFRKGYPVKISKTTFAQYRCGLPGDIAQCQNEKEIIDILQGGQVKYEDDF
KIGTEYVFLRERLAERYDGLQNKICGDAALIIQKNMKSFAQKVKRMRAAIKLOSG
LRGWKARRDYIIKREEMFKAIGRTMKRNKRLDAYHQALGTENSGQLQOTLVGYIDINE

gene

CDS

DAKFLERPTSDSATETLTHYLVPMKNFLSNMNSITLQFANEKGLLEPRREP
IMTPELHRESYDIFRLSVEIFKILKLYMNDIKLTKKQREDLGRIVQOGISNDCORDE
ILVQTINGKSNQDKTASDNGKMLVMAISVEPPTENIIPMLIGFNKESVPMKEQLF
ATLQRLKIIDSEIARELPSPNSLELIATPNIPNTVAELINCYDGLCCDVHLSPTSTSE
IAERILKRGVGIALGWTVEVERPNRVAFPTGNHFIHYESQIEGAMNDSENOQNIFF
NFPPEKLVKRPVVPKVEVQEOEIVSONTGTIPKARNTPTPEEMINOPQPTESRN
TPRDHQKDEESDESLEVRENEAGYATLPKRNKPAPQAVEDNETSESEDEDDAAN
RVGETVREPEENMKMETIODHSHILKSPVLPKRYTSPNEQHEEYTPMNFAPPEPTY
PQMPMAQYVPVMTSSMATPSMATPSMMPGOQIAYVPOQMMQPHFSYVPOYPOIPO
YCPPEILSPQSVRSEVPMPMAPVMDHGSTRSRDVRLMKRGEVPSQSTIRNMPV
EHGKDVDQFLDAVFDQVLSKDEORAAHFNSNOLANTIKGKLRPEGIYEPVQTYSPV
PPRYPTLRVDDSPLSRAKSLPRITISPRHEFVRRPHRSNSYNSTSSDDQMYRT
RSRERSLPRSSNGYNYDPSPQYMPMPVQMGHEMLILSPVASEHKAOSRRTDR
SHRHVPSGVERYMARRSPSVDLKPPMSRTPDAMVEDQTYVRPHLAOSPVGQTSRE
EEFSALPRGDSRAREMENPLARQIPRAYPYONGVYPPPPSSYRAPSPAPTSQDRR
GLNRLPQESTYEPVAVKNTKNNSEYLSPHRFNLDEQKAVIRHEKETAKNAIMNLSRLR
SPVQPTPPPPPPVPREEWVIRDSVERDPVTOGRIRGSRKEPLVPOPVRPVAEKP
AVKFKAPMKLTIRKEMFPGEVINDIOIIDQVFAQIYEDCKKSYPRIRREDRQVE
TVLROYEIPRPSDLNNSNHPDVAVIELARLMPLYFNQYEVVEKRPDESSTIEA
ISEHGIRLIVHTPHOLENPLKIDPFPEFETIADVSLSEANDILSVHRHEENAYSAV
RIKTNQAPQIKTKTLDRCLSGGVPRKRVLTARGRETTPGVNWLXKIENRFGLLAAQ
YVSTGDGNVPIRKHETSEDREVERFEDVEVPFSSERTMIDEATKYFRKPKDKKQ
ETWAMEDISQIVRSEKPIQSGLADLGNESKYAVETFAIMKMGDEPLKSESMT
DVVEFKVLLCHROPTLRDEYVQCOLIKROTSTINSOKPNSLARMLTITATVPSLSL
LKYFVLCYLDNADDEWQRPFGHTARICQTNMIQTFAEYGGKRVLLNALLEVQITVGS
TLVPECESTDCQQLRQAQFYISKDHANSQTLRPITVAEEMIQELCNLNVRSQHQOEE
SLCYTVGKRKHLNYCKNDNYLMDITTESEHKLPFOFYKRTVWVHPLRYDNAYIDS
MEDQVIDDYLRGSLISTNSLGQLAATAETEEIKLAAYLELLPDPNPKGLNAKTLPOIV
PKSVIEPKHROEEMVTRISRLKMGGRMPAEAKSHEPLESLTPTFGVHLRLKS
VVENGHQLPEVILITINKSGIQLLOPKSKEFKERNYDQIVSESIKTAIKYALVIN
TMOGEETLDIRKDEADEISHLIGOMFVTGGAERGSETE"

join(complement(3682..3795),complement(3460..3624),
complement(2750..3286),complement(2424..2700),
complement(2206..2376),complement(1899..2155),
complement(1748..1849),complement(1130..1425),
complement(1709..1079),complement(385..659),
complement(105..304),complement(268104.1:26738..27119),
complement(268104.1:26574..26660))
/gene="F46C3.1"
join(complement(3682..3795),complement(3460..3624),
complement(2750..3286),complement(2424..2700),
complement(2206..2376),complement(1899..2155),
complement(1748..1849),complement(1130..1425),
complement(1709..1079),complement(385..659),
complement(105..304),complement(268104.1:26738..27119),
complement(268104.1:26574..26660))
/note="similar to serine/threonine kinase (2 domains):
cDNA EST EMBL:D34010 comes from this gene; cDNA EST
EMBL:D36999 comes from this gene; cDNA EST EMBL:D64541
comes from this gene; cDNA EST yk459d11.3 comes from this
gene; cDNA EST yk459d11.3 comes from this gene; cDNA EST
yk459d11.5 comes from this gene; cDNA EST yk224a5.5 comes
from this gene; cDNA EST yk251c3.5 comes from this gene;
cDNA EST yk427c7.5 comes from this gene"

/codon_start=1
/protein_id="CAA91470.1"
/db_xref="PID:e1346871"
/db_xref="PID:g3877187"
/db_xref="GI:3877187"
/db_xref="SPTREMBL:Q19192"
/translation="MSVYIVLAGFLFMALVPFNAGQYIDDDIEVSSCQNGYVN
AGCDNSQVNIITATLDGVVTLADGETGEMIRYEDAPLRGLTSDPIDIGTSL
OLMPTLDGRLESTYHTNTNLEPLITDLSLESTIRLGDGAVAGKSVTKGFDLFTG
EOKYECSMSCGPGDTEPKNPILIRRTNSIRAMPDTGIERMNLSTAETGVLIA
GITSPTLYSDVKILLQPPDGVYAVADKYNREEMKTNVGHIVSVWQVYGNQIGETISIF
DPSNFTQYEVMEQEOHNLQSSLLYMGTSNGFPETIOSPKARNLKGORNALBEL
STMTELTNPRCTANETRSLAYNVKDETLRLVLHNAFRHSQSKAIEDKSLSGSSAR
KLQIIASDTEVSAORIGTENLRSSYSKSGDYGLVLESEPQRVKFVNSPTLMQTI
FSYIFNPTAVVSFLAGLIGVAVVAYNLIKAKSPRMIEHLSSTESAESESARHRTT
SFAPTDEIERFEVSGDSLSTPFIGAIHRKPLMPLEKSNIEHTHTQPIKPVORLVKTD
IDTDEDSFNSDEKKRLLRNRTISSRSLSEGTSRFANEFEVKKVYIGHGFGVFPRAQSI

CDS

gene

| | | | | |
|--------------------------|--------|----------------|-----------|---------------|
| Query Match | 12.68; | Score 40.4; | DB 36; | Length 29777; |
| Best Local Similarity | 50.08; | Pred. No. 3.5; | | |
| Matches 95; Conservative | 2; | Mismatches 93; | Indels 0; | Gaps 0; |

OY 43 ATGGAACAACAAAAAGTTTGTGCATTGTTAGTAGGAGTGCTGTAACCCCTTACT 102
 | | | | | | | | | | | | | |
Db 1473 ACAACAAAATAGATGCATCATTTGTTGAATTTTACGTGTTAATAATTAGCGAAACACA 1532

QY 103 CCCCATCAAAAAAAAAATGCATACATGCTTAAGGATARAAGGCCAATAATTTCATCATAT 162

Db 1533 GTCATTCAAAAATGCAAACGGCTTGAAGAAGCGGAATCGTTTTTGCGTTTTCGAA 15922

QY 163 GTCTAAAGAGAGAGAGAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGC 222

Db 1593 CGTATAGATCCGAAATGTGCTTAATAATAAAATATTTCAAAATGGAATTTCTTAAAAATTGC 1652

| | | | |
|----|------|-----------|------|
| QY | 223 | TTTGATAC | 232 |
| | | | |
| Db | 1653 | TTGGAATTG | 1662 |

RESULT 8

| LOCUS | AC007282 | 192151 bp | DNA | HTG | 05-JUN-1999 |
|------------|---|-----------|-----|-----|-------------|
| DEFINITION | Homo sapiens clone NH0484N09, WORKING DRAFT SEQUENCE, 4 unordered pieces. | | | | |

| | |
|-----------|-------------------|
| ACCESSION | AC007282 |
| NID | 95001487 |
| VERSION | AC007282.2 |
| KEYWORDS | GI:5001487 |
| | HTG; HTGS_PHASE1. |

REFERENCE
1 (bases 1 to 192151)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 192151)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 192151)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1999) Genome Sequencing Center, Washington

COMMENT

On Jun 5, 1999 this sequence version replaced gi:4580496. NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

| | | |
|--------|---------|-------------------------------|
| 1 | 5124: | contig of 5124 bp in length |
| 5125 | 5141: | gap of unknown length |
| 5142 | 32144: | contig of 27003 bp in length |
| 32145 | 32161: | gap of unknown length |
| 32162 | 105760: | contig of 73599 bp in length |
| 105761 | 105777: | gap of unknown length |
| 105778 | 192151: | contig of 86374 bp in length. |

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .192151 |

[illegible]

| | | | | |
|-----------------------|-----------------|----------------|-----------|----------------|
| Query Match | 12.5%; | Score 40; | DB 35; | Length 192151; |
| Best Local Similarity | 50.0%; | Pred. No. 3.6; | | |
| Matches 94; | Conservative 2; | Mismatches 92; | Indels 0; | Gaps 0 |

```
QY      29 TACAGTTACAACCAATAGGAAACAACAAAAGAAGTTGTGCATTGTAGTAGGAGCT 88  
        ||||| | | | | | | | | | | | | | | | | | | |  
Db 75422 TAGAGTAAAGACTCTGTCTCAAAAAAAAAAAAAAAGACTTATGGCAGCAGCTTAATATA 75481
```

QY 89 GTGTACCCCTTACTGCCCATCAAAAAAAAAATGGATACATGTTAAAGGATARAAGGCCA 148
| | | | | | | | | | | | : |
Db 75482 TTTAAGTCGACAGATTAAAGTTAAAGTAAGAATGTTACTACTACAGAAACCACCA 75541

QY 149 ATATTTTATCATATGTCTCTAAAGAGAGAGAGAGAAAATACTACTTTCTCRAAATGGAA 208
| | | | | | | | | | | | | | | | : | | | |
Db 75542 AACACTATTAATAATAGTAAAGGAGAGAAAGGAACAAAGTACACACAAAACACAGAAA 75601

| | | | |
|----|-------|----------|-------|
| QY | 209 | GGCCCTAA | 216 |
| | | | |
| Db | 75602 | TTAATTAA | 75609 |

RESULT 9
AF129078/c

| LOCUS | 78338 bp | DNA | PRI | 04-MAR-1999 |
|------------|--|-----|-----|-------------|
| DEFINITION | Homo sapiens chromosome 21q22.1 PAC B22238, complete sequence. | | | |
| ACCESSION | AF129078 | | | |

| | | |
|----------|------------|------------|
| VERSION | AF129078.1 | GI:4337089 |
| KEYWORDS | HTG. | |
| SOURCE | human. | |

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominini; Homidae; Homo.
REFERENCE 1 (bases 1 to 78338)

Rosenthal, A.
Direct Submission
Submitted (17-FEB-1999) Genome Analysis, Institute of Molecular
JOURNAL

JOURNAL Submitted (17-FEB-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany

FEATURES

source

Location/Qualifiers

1. .78338

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

3

```

repeat_region      complement(21..125)
                    /rpt_family="MT1A1"
                    /evidence=not_experimental
                    complement(332..407)
                    /note="MZEf, score = 83%"
                    /evidence=not_experimental
                    1315..1687
                    /rpt_family="THF1C"
                    /evidence=not_experimental
                    1817..1932
exon

```

```

repeat_region
/note="GRAIL, score = 59.000%, comment = good"
/evidence=not_experimental
1850..1928
/rpt_family="MLT1J"
/evidence=not_experimental
1860..1914
/note="Xpound exon prediction, score = 65% (0%)"
exon

```

```
repeat_region
/evidence=not_experimental
complement(1933..2079)
```

```

/ rpt_family="Alujo"
/evidence=not experimental

```

```
repeat_region
/evidence=not_experimental
2721. .3167
/rpt_family="LTR1"
```


exon complement(2762. .2844)
 /note="MZEf, score = 60.4%"
 /evidence=not_experimental
 repeat_region 3260. .3470
 /rpt_family="LTR28"
 /evidence=not_experimental
 complement(3471. .3538)
 /rpt_family="MT1A2"
 /evidence=not_experimental
 repeat_region 4631. .4700
 /note="homology = 70.00%, score = 21, counts = 10"
 /rpt_type=tandem
 /rpt_unit=aaaagaa
 /evidence=not_experimental
 complement(4825. .4872)
 /note="GRAIL, score = 84.000%, comment = excellent"
 /evidence=not_experimental
 exon 5249. .5414
 /note="Genscan, score = 4.31%, comment = Terminal_exon 166
 bp frame: 2 phase: 1"
 /evidence=not_experimental
 repeat_region 6153. .6187
 /rpt_family="L2"
 /evidence=not_experimental
 6600. .6701
 /rpt_family="L2"
 /evidence=not_experimental
 repeat_region 6723. .7190
 /rpt_family="MLT1H"
 /evidence=not_experimental
 6941. .6992
 /note="Xpound exon prediction, score = 66% (0%)"
 /evidence=not_experimental
 repeat_region 7367. .7665
 /rpt_family="L2"
 /evidence=not_experimental
 complement(9762. .9962)
 /rpt_family="MER3"
 /evidence=not_experimental
 repeat_region 9964. .10150
 /rpt_family="MIR"
 /evidence=not_experimental
 10362. .10639
 /rpt_family="AluSg"
 /evidence=not_experimental
 complement(10363. .10619)
 /note="MZEf, score = 81.5%"
 /evidence=not_experimental
 complement(10895. .11136)
 /rpt_family="MIR"
 /evidence=not_experimental
 repeat_region 13416. .13515
 /rpt_family="MER5A"
 /evidence=not_experimental
 14240. .14314
 /rpt_family="MER5A"
 /evidence=not_experimental
 14515. .14671
 /note="GRAIL, score = 95.000%, comment = excellent"
 /evidence=not_experimental
 exon 14515. .14671
 /note="MZEf, score = 56.7%"
 /evidence=not_experimental
 14766. .14791
 /note="111 with 96% homology to 112"
 /rpt_type=inverted
 /evidence=not_experimental
 complement(14808. .15551)
 /rpt_family="L1PA2"
 /evidence=not_experimental
 complement(14957. .15105)
 /note="GRAIL, score = 63.000%, comment = good shadow"
 /evidence=not_experimental

exon complement(15431. .15537)
 /note="GRAIL, score = 47.000%, comment = marginal"
 /evidence=not_experimental
 repeat_region 15552. .15577
 /note="112 with 96% homology to 111"
 /rpt_type=inverted
 /evidence=not_experimental
 repeat_region 16981. .17278
 /rpt_family="AluSg"
 /evidence=not_experimental
 complement(17569. .18083)
 /rpt_family="MER51B"
 /evidence=not_experimental
 complement(18364. .18887)
 /rpt_family="L1ME2"
 /evidence=not_experimental
 complement(18884. .19040)
 /rpt_family="L1ME2"
 /evidence=not_experimental
 complement(19041. .19170)
 /rpt_family="FLAM_C"
 /evidence=not_experimental
 19058. .19149
 /note="MZEf, score = 75.9%"
 /evidence=not_experimental
 complement(19171. .19315)
 /rpt_family="L1ME2"
 /evidence=not_experimental
 complement(19314. .19857)
 /rpt_family="L1MEc"
 /evidence=not_experimental
 complement(19837. .19944)
 /rpt_family="L1MEc"
 /evidence=not_experimental
 complement(19986. .20285)
 /rpt_family="AluSx"
 /evidence=not_experimental
 20834. .20904
 /note="MZEf, score = 52.3%"
 /evidence=not_experimental
 20834. .20904
 /note="GRAIL, score = 70.000%, comment = good shadow"
 /evidence=not_experimental
 complement(21748. .23646)
 /rpt_family="L1PA2"
 /evidence=not_experimental
 complement(21906. .22046)
 /note="GRAIL, score = 69.000%, comment = good shadow"
 /evidence=not_experimental
 23646. .23832
 /rpt_family="L1P"
 /evidence=not_experimental
 23833. .24258
 /rpt_family="LTR7"
 /evidence=not_experimental
 24267. .24324
 /rpt_family="L1P"
 /evidence=not_experimental
 24325. .24762
 /rpt_family="L1P"
 /evidence=not_experimental
 24341. .24410
 /note="MZEf, score = 87.1%"
 /evidence=not_experimental
 24763. .25071
 /rpt_family="AluJo"
 /evidence=not_experimental
 25072. .25617
 /rpt_family="L1P"
 /evidence=not_experimental
 25096. .25247
 /note="GRAIL, score = 71.000%, comment = good"
 /evidence=not_experimental

| Query Match | 12.4% | Score 39.8 | DB 42 | Length 78338 |
|-----------------------|--|---------------|----------|--------------|
| Best Local Similarity | 50.8% | Pred. No. 4.4 | | |
| Matches 92 | Conservative 1 | Mismatches 88 | Indels 0 | Gaps 0 |
| QY 31 | CAGTTACAACCAATAGGACACACAAAAGAAAAGTTGTGACATTGTAGTAGGAGTGT | 90 | | |
| DB 36167 | CAGGAAAAAAGAAAAAACAACAGAGGTTGACTATTAAAGCCAAATCAAGGAGGCACCAG | 36108 | | |
| QY 91 | GTACCCCTTACTCCCCATCAAAAAAATAATGATACATGTTAAAGATARAAGGCCAAT | 150 | | |
| DB 36107 | AGAAAGGTGAAGTCCCACTTAAGGAAAAAAGAAAAAGATATAGCATATGAATGGAAA | 36048 | | |
| QY 151 | ATTTTATCATATGTCTTAAAGAGAGAAGAGAAAATACTACTTCTCRAAATGGAAGC | 210 | | |
| DB 36047 | AATCTCAGTGAACACAGATAGCATTAATAAAAAACAATCACACTTCTCGGAAATCAAGGA | 35988 | | |
| QY 211 | C 211 | | | |
| DB 35987 | C 35987 | | | |

| | | | | |
|------------|--|------------|-----|-----------------|
| RESULT | 10 | | | |
| LOCUS | HS72E17/c | | | |
| DEFINITION | HS72E17 | 91029 bp | DNA | HTG 03-JUN-1999 |
| | Homo sapiens chromosome 6 clone 72E17, WORKING DRAFT SEQUENCE, in unordered pieces. | | | |
| ACCESSION | AL033523 | | | |
| NID | 95002632 | | | |
| VERSION | AL033523.5 | GI:5002632 | | |
| KEYWORDS | HTG; HTGS_PHASE1. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| REFERENCE | 1 (bases 1 to 91029) | | | |
| AUTHORS | Wlne,S. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humbrey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jun 7, 1999 this sequence version replaced qi:4678503. | | | |
| COMMENT | | | | |

```

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj72E17 Contig_ID: 01487 acc-AL033523
Length: 88938 bp Unfinished: dj72E17 Contig_ID: 01515
acc-AL033523 Length: 1291 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

| FEATURES | Location/Qualifiers |
|------------|---------------------------------|
| source | 1. .91029 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /chromosome="6" |
| | /clone="72E17" |
| BASE COUNT | 25175 a 20307 c 20181 g 24566 t |
| ORIGIN | 800 others |

| | | | | |
|-----------------------|--------------|----------------|-----------------|---------------|
| Query Match | 12.4%; | Score 39.6; | DB 34; | Length 91029; |
| Best Local Similarity | 49.0%; | Pred. No. 4.8; | | |
| Matches 99; | Conservative | 2; | Mismatches 101; | Indels 0; |
| | | | | Gaps 0; |

OY 7 AAGGCTTCCTTGTTAGAAATAGTACAGTTCAACCATTAGGAACAACAAAAAGT 66
| | | | | | | | | | | | | | | |
Db 44464 AGGTATTCCTATGTAAGATGAGCTCACAGACAAAAATTACGAGGCAGATAAGGAATTCAT 44405

| | | | |
|----|-------|--|-------|
| QY | 67 | TTGTGACATTTGTAGTAGGGAGTGTGTACCCCTTACTCCCCATCAAAAAAATGGAATAC | 126 |
| | | | |
| Db | 44404 | ATCAAAACACTGGTACAAACAATGTTAACTAAGTGCACAATAAATATGAGAAAAATATACAA | 44345 |
| QY | 127 | ATGGTTAAAGGATARAAGGGCAATATTTATCATATGTGTTCTTAAAGAGAAGGAAGAGAA | 186 |
| | | | |
| Db | 44344 | AAGAACCATGGATACACAGACCAATCAGAAATAAAGTTTAAACAAAGGAAGGAATAGAT | 44285 |
| QY | 187 | ATACTACTTTCTCRAAATGGAA | 208 |
| | | | |
| Db | 44284 | ATTAATGAACATAAAAAGGAAA | 44263 |

RESULT 11
PFMAL4P3/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

PFMAL4P3 312766 bp DNA HTG 09-APR-1999
Plasmodium falciparum chromosome 4 strain 3D7, WORKING DRAFT
SEQUENCE, in unordered pieces.
AL035476
94494052
AL035476.2 GI:4494052
HTG; HTGS_PHASE1.
malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 312766)
Bowman, S., Churcher, C., Harris, D., Lawson, D., Quail, M. and
Barrell, B.
Direct Submission
Submitted (09-APR-1999) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
On Mar 24, 1999 this sequence version replaced gi:4455722.

For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

| FEATURES | Location/Qualifiers |
|------------|--|
| source | 1. .312766 |
| | /organism="Plasmodium falciparum" |
| | /strain="3D7" |
| | /db_xref="taxon:5833" |
| | /chromosome="4" |
| BASE COUNT | 115941 a 29256 c 29590 g 116336 t 21643 others |
| ORIGIN | |

| | | | | |
|--------------------------|--------|----------------|-----------|----------------|
| Query Match | 12.4%; | Score 39.6; | DB 34; | Length 312766; |
| Best Local Similarity | 55.18; | Pred. No. 4.2; | | |
| Matches 75; Conservative | 1; | Mismatches 60; | Indels 0; | Gaps 0; |

OY 52 ACAAAAAGAAAAAGTTTGTGACATTGTAGTAGGGAGTGTTACCCCTTACTCCCACATCA 111
| | | | | | | | | | | | | | |
Db 94865 ACTAAGAAGACAAGATTTCCTTATATAATAATATGTCTATACCATATAGTCTTAATAAT 94806

OY 112 .AAAAAAATGCATACATGCTTAAAGCATARAAGCGCAATATTTTATCATATGTTCTAAA 171
||| ||| | ||| | | :|| | | | ||| | | | |||
Db 94805 AAACAAATTAAGAAATGCATGTAGACGAAATCAAAACATGTTATCCCTTTATCGAAA 947466

QY 172 GAGAGGAGAGAAA 187
| | | | | | | |
Db 94745 AATATGGAAGAGTCAA 94730

RESULT 12
AC007659/c 89603 bp DNA PLN 27-MAY-1999
LOCUS Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence,
DEFINITION complete sequence.
AC007659
AC007659
NID 94895233
VERSION AC007659.1 GI:4895233
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 89603)
Goodman,H.M., Agyare,F., Belmonte,S., Dolan,M., Kim,U. and
Morris,J.W.
Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence
Unpublished
2 (bases 1 to 89603)
Lin,X.
Direct Submission
Submitted (27-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
Address all correspondence to:
Howard Goodman
Mass. General Hospital
Boston MA 02114
goodman@mgm.harvard.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

This BAC was annotated in collaboration with TIGR group.
Contact:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone T14P1 is from Arabidopsis chromosome II and is near the
molecular marker m336.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/netgene/cbsnetgene.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.

FEATURES
Source
1. 89603
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="m336"
/clone="T14P1"
complement(1..539)
/gene="T14P1.1"

misc_feature
1..1868
/note="overlap with BAC clone F4L23
(AC002387:121061..122928)."
complement(1..539)
/gene="T14P1.1"
/note="F4L23..31: identical to GB:AAB82643"
complement(97..501)
/gene="T14P1.1"
/note="unknown protein"
/codon_start=1
/protein_id="AAD32819.1"
/db_xref="pid:g4895234"
/db_xref="gi:4895234"
/translation="MASKALAVTALLITNLFFFTVSTKCPPTPKPKPSRK
APAVKPTCPDTLKLGVGCADLGLVNVVGSPPKTPCCTLLQGLANLEAAVCLCTALK
ANVLGINLVNPIIDLTLNLNYCGKKVPHGQCS"
733..842
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(791..852)
/note="exon predicted by xgrail, quality marginal"
complement(882..898)
/note="exon predicted by xgrail, quality marginal"
complement(join(1180..1571,1671..1723,2116..2171,
2294..2341,2499..2594))
/gene="T14P1.2"
complement(1180..2594)
/gene="T14P1.2"
/note="F4L23..32: identical to GB:AAB82646. Two ESTs do not
appear to be completely spliced"
complement(join(1352..1571,1671..1723,2116..2171,
2294..2333))
/gene="T14P1.2"
/codon_start=1
/product="putative symbiosis-related protein"
/protein_id="AAD32820.1"
/db_xref="pid:g4895235"
/db_xref="gi:4895235"
/translation="MNKGSIFKMDDEKRAEAGRIRKYPPIPIVEKAEKSEVP
NIDKKYLVPSDLVGVGVYVIRKRIKLSAEKALFIFVDNVLPTGELMSSVYEDKKD
EDGFLYITYSGENTFGASSI"
complement(3021..3042)
/rpt_family="At_rich"
complement(3443..3571)
/note="exon predicted by xgrail, quality excellent"
3451..3563
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(4034..4074)
/note="exon predicted by xgrail, quality good"
complement(4109..4140)
/note="exon predicted by xgrail, quality marginal"
4194..4302
/rpt_family="(TAAA)n"
complement(4657..4723)
/note="exon predicted by xgrail, quality good"
complement(5290..5327)
/rpt_family="(TA)n"
complement(5891..5921)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
6344..6406
/note="exon predicted by xgrail, quality marginal"
complement(6361..6412)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
6589..6645
/rpt_family="(GAAA)n"
complement(6722..6814)
/rpt_family="(GAAA)n"
6845..8942
/gene="T14P1.3"
6845..8942
gene

CDS /gene="T14P1.3"
6845. .8767
/gene="T14P1.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAD32821.1"
/db_xref="PID:g4895236"
/db_xref="GI:4895236"
/translation="MPLSEFRFGEGVFGLSSSSFYSDSKIMSNODKTEAKQEDLGY
VVGGLPEPTSVLDALRSPSPASYSSTTTLSHGCGTNTVTAGDNNNNK
CSQMGDLDDGLVLSASSPGQESILRLIMDPGSAFGVDPGFGSGSPYSAVSDN
SNLCNFPFQETNPAPALINPANHCLFYNPLSPPAKFNSGSLHQPVEPLSDPDG
HDPVRQHQFOFPFYHNNQQQFPSSSSSTAVAMVPVSPGMAGDDQSVIIEDLFNAA
ELIGTGNNNGDHTVLAOGILARNHNLNTSSNHKSPQRAASHIAEALSLIHNESS
PPLITPENILIRIAAYRSFSESPFLQFVNFTANQSIESCNESEGFDRIHIIEDVG
GGOWSSLMQELASGVGRRRNRASSLKLTVEAPPSTVSDEFELRTEENLKTFAGEV
KIPFEIELLSVELLNPAYWPLSRSEKEAIAVNLPVNSVASGYLPILIRFLKQLSP
NIVVCSDRGCDRNDAPFPNAVHSLQYHTSLSDANQDDSSIERFWQDSIEKL
LMKRHRWIERSPFWRILFTQCGSPASLSQMAEAQAECLLRNPVRGFIHVKRQSSLY
MCWQRKELVTSAMKC"
repeat_region 10014. .10037
/rpt_family="AT_r1ch"
10070. .10118
/note="exon predicted by xgrail, quality good"
complement(10292. .10333)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
10433. .10467
/rpt_family="AT_r1ch"
10954. .11000
/rpt_family="(CAAAA)n"
complement(join(<11530. .11621,11704. .11819,11904. .12037,
12119. .12243,12366. .12556,12650. .12939,13131. .>13475))
/gene="T14P1.4"
complement(<11530. .>13475)
/gene="T14P1.4"
complement(join(11530. .11621,11704. .11819,11904. .12037,
12119. .12243,12366. .12556,12650. .12939,13131. .13475))
/gene="T14P1.4"
/codon_start=1
/product="putative phosphatidate cytidylyltransferase"
/protein_id="AAD32822.1"
/db_xref="PID:g4895237"
/db_xref="GI:4895237"
/translation="MTNTNTLFLRHDDSEILIFQENSVPFLVFASWGMIRAMATFA
ELVLSRGTCPKRSFTKRLPLIRPLSGRLPGDTKPLRGLRISVSRRELTAVAR
AESDQLGDDHSGKIDRIHNLNVEDKOKKASQKLKRYIFGIGLPGCVVLAGGW
FTVALASVFIQSREYELVRSRGLAKMTPPRYRVSVCICALMPILTYEKNID
ILVTSAPVVAIALLVORGSPPRAQLSSTMGLFYCGYLPFWKLRGLADPALNTG
IGRTWPIILGGQAHMTVGLVATLIFSGLVIATDTFAFLGKTEGRTPLTSISPKTWE
GRIVGLVGIATITILSKYLSWPQSLFSSVAFGLNFGSVFEGDLTESMIKRDAGVKD
SGSLPGHGAILDRVDSYIFGTALAYSFIKTSKLKYG"
complement(join(13660. .14019,14258. .14335,14868. .14918,
14999. .15157,15239. .15354,15460. .15541,15621. .15691,
15800. .15877))
/gene="T14P1.5"

Query Match 12.1%; Score 38.8; DB 8; Length 89603;
Best Local Similarity 48.5%; Pred. No. 7.2;
Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0;

QY 33 GTTACAACCAATAGAACACAAAAGTTGTGACATTGTAGTAGGAGTGTGT 92
Db 84605 GTTTGATCATACACAATGTAGAACAAAAGAGATTATTTGAAAATCAATTGCGG 84546
QY 93 ACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGGATARAAGGCAATAT 152
Db 84545 AAACCATGTTCCAATTAACAACAAAAGAGAGTATCACCAAGACATAAACCCCTAACAG 84486
QY 153 TTATCATATGTCTAAAGAGAAGAGAAATACTACTTTCTCRAAATGGAAGCCC 212
Db 84485 TGTACCTATTCTCTATCATCAAAATCCCAAAATGTTCTCTCAAAAACCTGATT 84426

QY 213 TTAAGGTGCTTTGATCTGAAGAC 238
Db 84425 TGCCATATCCAAATTATCTTCATGAC 84400
RESULT 13
ATAC002388/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
ATAC002388 85534 bp DNA PLN 14-DEC-1998
Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence,
complete sequence.
AC002388
93420042
AC002388.1 GI:3420042
HTG.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 85534)
Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R.
and Venter,J.C.
Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence
Unpublished
2 (bases 1 to 85534)
Rounsley,S.D. and Lin,X.
Direct Submission
Submitted (25-JUL-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, Email:
rounsley@tigr.org
3 (bases 1 to 85534)
Rounsley,S.D.
Direct Submission
Submitted (27-AUG-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 85534)
Rounsley,S.D.
Direct Submission
Submitted (15-AUG-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Aug 15, 1998 this sequence version replaced gi:2344885.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr.,
Rockville, MD 20850,
USA
e-mail: rounsley@tigr.org
BAC clone T13E15 is from Arabidopsis chromosome II and contains the
molecular marker m336.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
as isologs. Genes without significant peptide similarity but with
EST similarity are named as 'unknown' proteins. Genes without
protein or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
'hypothetical' proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by genome.wustl.edu/eddy/tRNAscan-SE/).
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted

| FEATURES | | exons by GRALL are annotated as misc features. |
|---------------|--|--|
| source | Location/Qualifiers | |
| CDS | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m36" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLDDVDSSDAGSVNVEFGDCTNPSTVSHEPAIIFS CYDSDGNMGRGMEFALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQNTQA SKPLWAVAVTQSYNSRRKVRSSISIPLESCLAKASFASQKSASLEMPRIGYQD SDRSQWTVTERLKRKSSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| mRNA | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGSRDSESVTPLVGGGNGDNVAVAPQVENSPLALNEASYI TQATSYLSLEYGKDTISHPHELRSTSGVDGNSPVSCISPERSTSEASTSDA NSPSRESTEILPQATNAIVTSNRNLNGISMFOGLIERARTVRGSDADDIGLQRAPE MPVEDGTDRFNKILEDIGNHGPLYVDTKRFSKMGGLACHIAKIHSSSVEKNAREI KEYIEELCWGSKRVLLIGHSGKIDAAALSLYPELKDVAAGLVAAQSPYGGPIA TDILREGOLDVYNLRKMEILISKVIKGDQALDLYERKRKEFLKHPLEPLPTV SFRTEASISPAVLSTLSHVAHAELPLTNQAAKLPEVMPLGAAACQDLLOVRYGEKS DELVTCDDAEVPGSVVRRKRKLDAHMMVSSLINEVPLEADAQVCEALLTLVQVEQ ERQOKLATRKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrall, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grall" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRHSSISPSSEAAAVKQPFEPSSSEDKLDTAV LQFONLKLQKLEAQVCESSILEDKLSQIKEKQLPYNSLKTVMKSWEQDLTASVESC SVRVSDSSSGAHRFVNKEDSSPAVKNDFINRLLETGATESSSSNICSNOMEENGVT SSQMTQTLVNLVAATEDLRCILKDELPTVLRNLGKDLGOLALSELSEIKSFRGDL DDVLVFKSLSLRELQSHRADAKAVRVDLKRIRGSILVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrall, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| FEATURES | | exons by GRALL are annotated as misc features. |
|---------------|--|--|
| source | Location/Qualifiers | |
| CDS | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m36" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLDDVDSSDAGSVNVEFGDCTNPSTVSHEPAIIFS CYDSDGNMGRGMEFALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQNTQA SKPLWAVAVTQSYNSRRKVRSSISIPLESCLAKASFASQKSASLEMPRIGYQD SDRSQWTVTERLKRKSSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| mRNA | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGSRDSESVTPLVGGGNGDNVAVAPQVENSPLALNEASYI TQATSYLSLEYGKDTISHPHELRSTSGVDGNSPVSCISPERSTSEASTSDA NSPSRESTEILPQATNAIVTSNRNLNGISMFOGLIERARTVRGSDADDIGLQRAPE MPVEDGTDRFNKILEDIGNHGPLYVDTKRFSKMGGLACHIAKIHSSSVEKNAREI KEYIEELCWGSKRVLLIGHSGKIDAAALSLYPELKDVAAGLVAAQSPYGGPIA TDILREGOLDVYNLRKMEILISKVIKGDQALDLYERKRKEFLKHPLEPLPTV SFRTEASISPAVLSTLSHVAHAELPLTNQAAKLPEVMPLGAAACQDLLOVRYGEKS DELVTCDDAEVPGSVVRRKRKLDAHMMVSSLINEVPLEADAQVCEALLTLVQVEQ ERQOKLATRKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrall, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grall" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRHSSISPSSEAAAVKQPFEPSSSEDKLDTAV LQFONLKLQKLEAQVCESSILEDKLSQIKEKQLPYNSLKTVMKSWEQDLTASVESC SVRVSDSSSGAHRFVNKEDSSPAVKNDFINRLLETGATESSSSNICSNOMEENGVT SSQMTQTLVNLVAATEDLRCILKDELPTVLRNLGKDLGOLALSELSEIKSFRGDL DDVLVFKSLSLRELQSHRADAKAVRVDLKRIRGSILVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrall, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| | | |
|--|--|--|
| Query Match | | 12.1%; Score 38.8; DB 8; Length 85534; |
| Best Local Similarity | | 48.5%; Pred. No. 7.2; |
| Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0; | | |
| Db | | 11128 AACCATGTTCCAAATATACAAACAAAGAGATATACCAAGACATAAACCTAACAG 11069 |
| QY | | 33 GTTACACCAATAGACACAAAGAAAGTTGTGACATGTGTAGAGAGTGTGT 92 |
| DB | | 11188 GTTGGATCATACACAATGTAGACAAAGAAAGAGATTAATTGAAAAATCAATTGCGG 11129 |
| QY | | 93 ACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGATARAAGGCCAATAT 152 |
| DB | | 11128 AACCATGTTCCAAATATACAAACAAAGAGATATACCAAGACATAAACCTAACAG 11069 |


```
repeat_region      21434. .21497
                    /rpt_family="(CAAAA)n"
repeat_region      21719. .21743
                    /rpt_family="(TAAA)n"
repeat_region      complement(21747. .21869)
                    /rpt_family="AT_rich"
repeat_region      complement(22138. .22846)
                    /rpt_family="Lx4"
repeat_region      complement(22875. .23251)
                    /rpt_family="L1_MM"
repeat_region      complement(23235. .23520)
                    /rpt_family="Lx4"
repeat_region      complement(23520. .23698)
                    /rpt_family="L1"
repeat_region      complement(23926. .24056)
                    /rpt_family="Lx3"
repeat_region      complement(24353. .24390)
                    /rpt_family="AT_rich"
repeat_region      24885. .25275
                    /rpt_family="Lx4"
repeat_region      complement(25276. .25307)
                    /rpt_family="POLY_A"
repeat_region      complement(25416. .26556)
                    /rpt_family="L1_MM"
repeat_region      26657. .28391
                    /rpt_family="Lx4"
repeat_region      complement(28448. .28491)
                    /rpt_family="AT_rich"
repeat_region      28945. .29032
                    /rpt_family="AT_rich"
repeat_region      complement(29898. .29954)
                    /rpt_family="(TAAA)n"
repeat_region      complement(29957. .30863)
                    /rpt_family="Lx2B"
repeat_region      complement(30905. .32970)
                    /rpt_family="L1"
repeat_region      complement(33018. .33876)
                    /rpt_family="L1"
repeat_region      complement(33895. .33998)
                    /rpt_family="L1"
repeat_region      complement(34617. .34657)
                    /rpt_family="AT_rich"
repeat_region      35012. .35139
                    /rpt_family="B1_MM"
repeat_region      complement(35253. .36201)
                    /rpt_family="L1"
repeat_region      complement(36211. .37107)
                    /rpt_family="Lx2B"
repeat_region      37094. .37414
                    /rpt_family="Lx3"
repeat_region      37417. .37598
                    /rpt_family="Lx2"
repeat_region      complement(37613. .38174)
                    /rpt_family="L1VL2"
repeat_region      complement(38118. .38381)
                    /rpt_family="L1VL2"
repeat_region      38410. .38945
                    /rpt_family="Lx7"
repeat_region      complement(39080. .39151)
                    /rpt_family="AT_rich"
repeat_region      39216. .39249
                    /rpt_family="AT_rich"
repeat_region      complement(39401. .39940)
                    /rpt_family="RMR6A"
repeat_region      41499. .41933
                    /rpt_family="L1_MM"
```

```
Query Match      12.1%; Score 38.8; DB 12; Length 69037;
Best Local Similarity 48.5%; Pred.No.7.4;
Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0;

QY      11 CTCTCTGTAGAAATAGTACAGTTACACCAATAGACACAAAAAGAAAGTTTGT 70
        ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 65351 CTCTATTAAATTAACACCAATTTTAACATGATAGAAAAAAATCTAGCCACATGTGG 65410
QY      71 GACATTGTAGAGAGTGTTGTACCCTTACTCCCATCAAAAAAAATGATACATGG 130
        | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65411 TATAGAGAGTCAACATAAAAAATCCCTAGGGGACTGAAAAAAATATAATTATGATGT 65470
QY      131 TTAAGATARAAGGCCATATTTATTCATATGTCTTAAAGAGAAAGAGAAAAATAC 190
        ||||| ||: || | | | | | | | | | | | | | | | | | | | |
Db 65471 ATAACATTAGGAGCAGAAATATTTTACTCATTTTCATCTTAAAGATGACTAATGTCC 65530
QY      191 TACTTCTCRAAATGGAAGCCCTAA 216
        || ||| : || | | | | | | | | | | | | | | | | | | | |
Db 65531 TATGCTCTTAAACATTAAGCCAATTA 65556
```

```
RESULT 15
HS800F24/c
LOCUS      HS800F24      138888 bp      DNA      PRI      12-MAY-1999
DEFINITION Human DNA sequence from clone 800F24 on chromosome 1q24, complete
sequence.
ACCESSION      299758
NID      94775628
VERSION      299758.7      GI:4775628
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE      1 (bases 1 to 138888)
JOURNAL      Grafham,D.
COMMENT      Direct Submission
Submitted (12-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On May 11, 1999 this sequence version replaced gi:4678515.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL
IMPORTANT: This sequence is not the entire insert of clone 800F24.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone dj1018K9 is at 138789 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
800F24 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pcypac2.
```

```
FEATURES
source
1. .138888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="800F24"
/map="q24"
/clone_1tb="RPC14"
1. .1964
/note="L1 repeat: matches 3306. .5335 of consensus"
repeat_region      1967. .2267
```


| | | | |
|---------------|--|---------------|---|
| repeat_region | /note="AluSc repeat: matches 1. .300 of consensus" | repeat_region | 27202. .27345 |
| repeat_region | /note="LIP2 repeat: matches 5413. .5484 of consensus" | repeat_region | /note="L2 repeat: matches 2559. .2710 of consensus" |
| repeat_region | /note="LIM1 repeat: matches 799. .1710 of consensus" | repeat_region | 27849. .28052 |
| repeat_region | /note="LIM1 repeat: matches 491. .794 of consensus" | repeat_region | /note="MIR repeat: matches 1. .205 of consensus" |
| repeat_region | /note="MIR repeat: matches 87. .216 of consensus" | repeat_region | 28914. .29215 |
| repeat_region | /note="MIR repeat: matches 49. .146 of consensus" | repeat_region | /note="AluSc repeat: matches 1. .310 of consensus" |
| repeat_region | /note="LTR16B repeat: matches 1. .464 of consensus" | repeat_region | 29892. .30582 |
| repeat_region | /note="AluSq repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP16 repeat: matches 5476. .6155 of consensus" |
| repeat_region | /note="L2 repeat: matches 2613. .2710 of consensus" | repeat_region | 30596. .31979 |
| repeat_region | /note="MIR repeat: matches 200. .250 of consensus" | repeat_region | /note="LIP16 repeat: matches 4744. .6155 of consensus" |
| repeat_region | /note="AluDb repeat: matches 85. .287 of consensus" | repeat_region | 32709. .32764 |
| repeat_region | /note="MIR repeat: matches 72. .116 of consensus" | repeat_region | /note="L2 repeat: matches 2654. .2709 of consensus" |
| repeat_region | /note="MIR repeat: matches 72. .116 of consensus" | repeat_region | 34717. .35145 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 1906. .2371 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 35142. .35391 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 2141. .2359 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 35392. .35691 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="AluSc repeat: matches 1. .300 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 35692. .37115 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 477. .2141 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 37415. .37828 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MIR repeat: matches 6. .426 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 38428. .38572 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MIR repeat: matches 111. .262 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 38626. .39246 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIM4 repeat: matches 4297. .4933 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 39346. .39651 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="AluDo repeat: matches 1. .301 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 39699. .40087 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 1879. .2298 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 40158. .40665 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIM4 repeat: matches 4796. .5332 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 40666. .40992 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="TIGER1 repeat: matches 2078. .2418 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 41012. .42354 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIP16 repeat: matches 4784. .6144 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 43109. .43233 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MIR repeat: matches 20. .148 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 43333. .43694 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 1188. .1584 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 44121. .44349 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MER8 repeat: matches 8. .239 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 44360. .44465 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 2602. .2712 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 44502. .44695 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="L2 repeat: matches 2302. .2491 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 45695. .45961 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="AluDo repeat: matches 1. .299 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 46075. .46116 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="21 copies 2 mer at 79% conserved" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 46137. .46200 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="32 copies 2 mer tt 67% conserved" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 46646. .46790 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIME repeat: matches 5607. .5751 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 46899. .47049 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MIR repeat: matches 102. .258 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 47326. .47575 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="MER46A repeat: matches 1. .236 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 47845. .48314 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIP repeat: matches 4916. .5395 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 48303. .52827 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIP16 repeat: matches 1236. .5781 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 52826. .57156 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIP16 repeat: matches 2. .4320 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 57157. .57376 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="AluV repeat: matches 94. .311 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 57377. .59193 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="LIP16 repeat: matches 4320. .6144 of consensus" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 58567. .58667 |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | /note="Single clone region" |
| repeat_region | /note="LIMEC repeat: matches 2371. .2296 of consensus" | repeat_region | 59220. .59596 |

Query Match

12.0%; Score 38.4; DB 10; Length 138888;

| | | |
|-----------------------|------------------|----------------|
| Best Local Similarity | 50.0%; | Pred. No. 8.4; |
| Matches | 93; Conservative | 1; Mismatches |
| | | 92; Indels |
| | | 0; Gaps |
| | | 0; |

20 AGAAATAGTACGTTACAACCAATAGGACACAAAAAGTTGTGACATTGTA 79

Db 125120 AGAAGGAAAAAGAGACCAACAAAGAAAAACAAAACAAATGATTTTTCCT 125061

80 GTAGGAGTGTGTACCCCTTACTCCCCCATCAAAAAAAAAATGGATACATGCTTAAAGGAT 139

Db 125060 TTAAGGCAATCTTTTCACCTTATACTAGAAATAGTCTGAAAAATGCAAGTAT 125001

140 ARAAGGGCATATTTTATCATATGTTCTAAAGAGAAGGAGAGAAATACTACTTTC 199

Db 125000 TTATGAGTTTGGTGATCAGTATGAATAAGCTTTAATAACAGGACAATGTTACTTCTA 124941

QY 200 RAAATG 205

Db 124940 A C A T T G 124935

Search completed: September 28, 1999, 12:43:44
Job time: 6304 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:30:02 ; Search time 289.74 Seconds
(without alignments)
276.322 Million cell updates/sec

| | |
|----------------|---|
| Title: | US-09-030-606-224 |
| Perfect score: | 320 |
| Sequence: | 1 CCCCTGAAGCCTCTGTGTA.....TTTAACTCMGCAATGTGAC 320 |

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| 1 | 318 | 99.4 | 320 | 1 | V58694 | Prostate tumour sp |
| 2 | 318 | 99.4 | 320 | 1 | V61269 | CDNA sequence of p |
| 3 | 45 | 14.1 | 110000 | 1 | V21209_01 | Continuation (2 of |
| 4 | 37 | 11.6 | 110000 | 1 | V21209_15 | Continuation (16 o |
| 5 | 36.6 | 11.4 | 19124 | 1 | T72882 | Plasmodium var-7 g |
| 6 | 35.6 | 11.1 | 1215 | 1 | T68041 | H. pylori cytoplasm |
| 7 | 35.6 | 11.1 | 717 | 1 | T67711 | H. pylori cytoplasm |
| 8 | 35 | 10.9 | 1107 | 1 | V74949 | Staphylococcus aur |
| 9 | 34.6 | 10.8 | 1811 | 1 | V17183 | Insecticidal toxin |
| 10 | 34.2 | 10.7 | 10813 | 1 | V74675 | Staphylococcus aur |
| 11 | 34.2 | 10.7 | 110000 | 1 | X20248_04 | Continuation (5 of |
| 12 | 34.2 | 10.7 | 116277 | 1 | X20249 | Continuation (5 of |
| 13 | 34 | 10.6 | 9842 | 1 | X20259 | Borrelia burgdorfer |
| 14 | 33.8 | 10.6 | 110000 | 1 | X20248_03 | Borrelia burgdorfer |
| 15 | 33.2 | 10.4 | 7295 | 1 | T72883 | Continuation (4 of |
| 16 | 32.8 | 10.2 | 64976 | 1 | V21209_16 | Plasmodium var-1 g |
| 17 | 32.8 | 10.2 | 2120 | 1 | X33812 | Continuation (17 o |
| 18 | 32.6 | 10.2 | 110000 | 1 | X20248_08 | Coding sequence fo |
| 19 | 32.4 | 10.1 | 1406 | 1 | T41619 | Continuation (9 of |
| 20 | 32.4 | 10.1 | 14736 | 1 | V52304 | Tomato endo-xylogl |
| 21 | 32.2 | 10.1 | 400 | 1 | V87584 | Streptococcus pneu |
| 22 | 32 | 10.0 | 4093 | 1 | Q49264 | EST clone DX307. N |
| 23 | 32 | 10.0 | 1457 | 1 | V15409 | ced-4. Long-distanc |
| 24 | 32 | 10.0 | 80595 | 1 | V83939 | Methicillin resist |
| 25 | 32 | 10.0 | 80240 | 1 | V83940 | HC-contig derived |
| 26 | 32 | 10.0 | 18359 | 1 | X20255 | NC-contig derived |
| 27 | 31.8 | 9.9 | 1920 | 1 | Q88066 | Borrelia burgdorfer |
| 28 | 31.8 | 9.9 | 1920 | 1 | Q95031 | Bacillus lichenifo |
| 29 | 31.8 | 9.9 | 1920 | 1 | V02471 | Bacillus lichenifo |
| 30 | 31.8 | 9.9 | 110000 | 1 | V21209_10 | Continuation (11 o |
| 31 | 31.6 | 9.9 | 134525 | 1 | Q04525 | Total base sequenc |
| 32 | 31.6 | 9.9 | 5852 | 1 | Q11710 | Dictyostelium plas |
| 33 | 31.6 | 9.9 | 5596 | 1 | V83941 | Bacterial artificia |
| 34 | 31.4 | 9.8 | 2207 | 1 | N80748 | Sequence of a poly |
| 35 | 31.4 | 9.8 | 1753 | 1 | V18455 | Human granulocytic |
| 36 | 31.4 | 9.8 | 110000 | 1 | X20248_01 | Continuation (2 of |
| 37 | 31.4 | 9.8 | 110000 | 1 | X20248_02 | Continuation (3 of |
| 38 | 31.4 | 9.8 | 10715 | 1 | X20248_09 | Continuation (10 o |
| 39 | 31.2 | 9.8 | 4472 | 1 | T05406 | Human cell cycle g |
| 40 | 31 | 9.7 | 4059 | 1 | T08973 | CI gene. Transgeni |
| 41 | 31 | 9.7 | 110000 | 1 | V21209_13 | Continuation (14 o |
| 42 | 31 | 9.7 | 110000 | 1 | V21209_14 | Continuation (15 o |
| 43 | 31 | 9.7 | 2025 | 1 | V54207 | Nucleotide sequenc |

ALIGNMENTS

[illegible]

| | | | | |
|---------------------------|---------|--------------------|-----------|-------------|
| Query Match | 99.4%; | Score 318; | DB 1; | Length 320; |
| Best Local Similarity | 100.0%; | Pred. No. 1.3e-74; | | |
| Matches 320; Conservative | 0; | Mismatches 0; | Indels 0; | Gaps 0 |

| | | | | |
|--------|--------------------------------|-----|--|-----|
| OY | | 1 | CCCCGAAAGCCTTCTGTGGAAAAATAGTAGACGTTACAACCAATTAGGAACAAACAAAAGA | 60 |
| Dp | | 1 | CCCGTGAAAGGCCTCTGTGGAAAAATAGTAGACGTTACAACCAATTAGGAACAAACAAAAGA | 60 |
| OY | | 61 | AAAAGTTGTGACATTTGTAGTAGGAGTGTGTACCCTTACTCCCACATAAAAAAAAAAAT | 120 |
| Dp | | 61 | AAAAGTTGTGACATTTGTAGTAGGAGTGTGTACCCTTACTCCCACATAAAAAAAAAAAT | 120 |
| OY | | 121 | GATACATGGTTAAAGGATARAAGGCCAATATTTATCATATGTTCTAAAGAGAAGAA | 180 |
| Dp | | 121 | GATACATGGTTAAAGGATARAAGGCCAATATTTATCATATGTTCTAAAGAGAAGAA | 180 |
| OY | | 181 | GAGAAAAATACTACTTTCTCRRAATGGAAGCCCTTAAGGTGCTTGTACTGAAGGACAC | 240 |
| Dp | | 181 | GAGAAAAATACTACTTTCTCRRAATGGAAGCCCTTAAGGTGCTTGTACTGAAGGACAC | 240 |
| OY | | 241 | AATGTGGCCGTTCATCCTCTTTARAGTTGCATGACTTGACACGGTAACGTGGCAGT | 300 |
| Dp | | 241 | AATGTGGCCGTTCATCCTCTTTARAGTTGCATGACTTGACACGGTAACGTGGCAGT | 300 |
| OY | | 301 | TATARACTCMGCATTTGTGAC | 320 |
| Dp | | 301 | TATARACTCMGCATTTGTGAC | 320 |
| RESULT | 2 | | | |
| ID | V61269 standard; cDNA; 320 BP. | | | |
| AC | V61269; | | | |

DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P510S.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 124; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 320 BP; 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAAAGA 60
Db 1 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAAAGA 60
OY 61 AAAAGTTGTGACATGTAGTAGGAGTGTGTACCCCTTACTCCCACTCAAAAAAAT 120
Db 61 AAAAGTTGTGACATGTAGTAGGAGTGTGTACCCCTTACTCCCACTCAAAAAAAT 120
OY 121 GGATACATGTTAAAGGATARAAGGCAATTTTATCATATGTCTTAAAGAGAGAA 180
Db 121 GGATACATGTTAAAGGATARAAGGCAATTTTATCATATGTCTTAAAGAGAGAA 180
OY 181 GAGAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATGTAAGGACAC 240
Db 181 GAGAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATGTAAGGACAC 240
OY 241 AAATGTGGCCGTCACCTCTTTARAGTTGCATGATGACACGCTAAGTGTGACAGT 300
Db 241 AAATGTGGCCGTCACCTCTTTARAGTTGCATGATGACACGCTAAGTGTGACAGT 300
OY 301 TTARACTCMGCATGTGAC 320
Db 301 TTARACTCMGCATGTGAC 320

RESULT 3

V21209_01/c Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular chr

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |

WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match 14.1%; Score 45; DB 1; Length 110000;
Best Local Similarity 48.2%; Pred. No. 0.0093;
Matches 120; Conservative 2; Mismatches 127; Indels 0; Gaps 0;

OY 5 TGAAGCTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAAAGAAAA 64
Db 93590 TGAAGACACCTAAAGATGAGAAATATTTAAACCAATATCCAAACATATATAAAA 93531
OY 65 GTTGTGACATGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGAT 124
Db 93530 ATTGTGAAGTTTAAAGACAGAAATGGAGATAGTACCCTTCCAAAAAATGGTT 93471
OY 125 ACATGTTAAAGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAGAGA 184
Db 93470 TAATTGGCAGGCAATGAGAAACTATTAATTTTAAATAAAAAATCTAAATAAAAACGAAT 93411
OY 185 AATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATGTAAGGACACAAAT 244
Db 93410 ATATGTTCCAACCTTAGACAGAAAAACCATATAATAGTTTGTAGAGATGATGAAT 93351
OY 245 GTGGCCGTC 253
Db 93350 ATTACCTTC 93342

RESULT 4

V21209_15/c Continuation (16 of 17) of V21209 from base 150001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

Query Match 11.6%; Score 37; DB 1; Length 110000;
Best Local Similarity 49.2%; Pred. No. 1.1;
Matches 94; Conservative 1; Mismatches 96; Indels 0; Gaps 0;

OY 56 AAAGAAAAGTTGTGACATGTAGTAGGAGTGTATCCCTTACTCCCATCAAAAA 115
Db 61937 AAAGAAAAGTTTAAATACTAAAGGGTTTATATTATGAGTAGTACTTACCCTTAGAAA 61878
OY 116 AAATGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGA 175
Db 61877 ATATGATAGAAAAGCTTAATATTTGGCGAGTTGAAGTATATTATGTTGAATGATT 61818
OY 176 AGGAAGAAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATAGTGAAG 235
Db 61817 GCCCTATTAAATAAGACCGTTTCGGAATGGAATCGTAATGCTTCTTATGACT 61758
OY 236 GACACAATGT 246
Db 61757 AAAAACACAGT 61747

RESULT 5
ID T72882 standard; cDNA; 19124 BP.
AC T72882;
DT 12-SEP-1997 (first entry)
DE Plasmodium var-7 gene.
KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
OS Plasmodium vivax.
OS Plasmodium falciparum.
FH Key location/Qualifiers
FT exon 7317. .15139
FT /*tag= a
FT /number= 1
FT 15140. .16205
FT /*tag= b
FT /number= 1
FT 16206. .17552
FT /*tag= c
FT /number= 2
FT /note= "no stop codon given"
PN WO9640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09508.
PR 07-JUN-1995; US-487826.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
DR WPI; 97-052231/05.
DR P-PSDB; W22475.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and stalic acid
PT binding proteins
PS Claim 4; Page 56-61; 96pp; English.
CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and stalic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 11.4%; Score 36.6; DB 1; Length 19124;
Best Local Similarity 51.5%; Pred. No. 0.96;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
OY 36 ACAACCAATAGGAACAACAAAAAGTTGTGACATTGTAGTAGGAGTGTACC 95
DB 402 ACNAAAAAAAAAAAAAAAAAAAAAAAAATTATATATAATATAAAAAAAAAAAGA 461
OY 96 CCTTACTCCCATCAAAAAAAAAATGCATACATGTTAAAGCATARAAGGCAATATTTT 155
DB 462 CATACGTTCACTTATTATTATAATGATTATTACGATTAAACATATTTAGATTATAAT 521
OY 156 ATCATATGTTCTAAAGAGAAGAGAAGAAATACTACTTCT 198

DB 522 AATATATTTTACATAGAAAGAGTTAAGAATACATTTT 564
RESULT 6
ID T68041 standard; DNA; 1215 BP.
AC T68041;
DT 21-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 07gp11807orf42.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS Helicobacter pylori.
FH Key location/Qualifiers
FT cds 1. .1215
FT /*tag= a
FT /note= "no stop codon given"
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Møllgaard BL;
DR WPI; 97-052306/05.
DR P-PSDB; W20788.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 9; Page 862; 1481pp; English.
CC This sequence encodes a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 1215 BP; 379 A; 254 C; 214 G; 368 T;

Query Match 11.1%; Score 35.6; DB 1; Length 1215;
Best Local Similarity 56.0%; Pred. No. 0.93;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;
OY 128 TGGTTAAGGATARAAGGCCATATTTATCATATGTTCTAAAGAGAAGAGAAAA 187
DB 857 TGGGTAATGGGCAAGCCCTGATTTATGTTAGATTTTAAAGAAATGAAGAAAAAT 916
OY 188 TACTACTTCTCRAAATGCAAGCCCTTAAGGTCCTTGATCTGAAGACACAAA 243
DB 917 TCATCATTTTTCATGCGGAACCGCTTAAAGCCTTAAAGCGCATGCGTTAAA 972

RESULT 7
ID T67711 standard; DNA; 717 BP.
AC T67711;
DT 17-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 719606.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS Helicobacter pylori.
FH Key location/Qualifiers
FT cds 1. .717

FT /tag- a
FT /note- "no stop codon given"
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR P-PSDB; W20567.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 9; Page 315-318; 1481pp; English.
CC This sequence encodes a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 717 BP; 224 A; 127 C; 146 G; 220 T;

Query Match 11.1%; Score 35.6; DB 1; Length 717;
Best Local Similarity 56.0%; Pred. No. 0.82;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

OY 128 TGGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAAGAGAAAA 187
| | | | | : | | | | | | | | | | | | | | | | | | | | |
Db 434 TGGTTAATGGCCAGCCTGATTATTGTTAGATTTTAAAGAAATGAAGAAAAAT 493
| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 188 TACTACTTCTCRAAATGGAAGCCCTTAAAGCTGCTTTGATACCTGAACACAAA 243
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 494 TCATCATTTTTCATGCGGAACCGCTTAAAGCCTTTAAAGCGCATGGCGTTAA 549

RESULT 8
V74949
ID V74949 standard; DNA; 1107 BP.
AC V74949;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #638.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 541..600
FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines
PS Claim 1; Page 1565-1566; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 1107 BP; 384 A; 134 C; 194 G; 329 T;

Query Match 10.9%; Score 35; DB 1; Length 1107;
Best Local Similarity 57.7%; Pred. No. 1.3;
Matches 56; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

OY 108 TCAAAAAAATGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCT 167
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 221 TAAAAATAAAAATTGGGGAGAGTTATATGCTAAACATTCAGACGTTANCATNCTTCT 280
| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 168 AAAAGAGAGAGAGAGAAATACTTCTCTCRAAT 204
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 281 AAAAAGAGCAAAAAGCATATACCGTTTCGTAGAAT 317

RESULT 9
V17183
ID V17183 standard; DNA; 1811 BP.
AC V17183;
DT 18-JUN-1998 (first entry)
DE Insecticidal toxin subunit BrhTX-1(a)1.1 genomic sequence.
KW Insecticidal toxin; Bracon hebetor; insect control; pathogen;
KW recombinant baculovirus; ds.
OS Bracon hebetor.
PN W09744355-A1.
PD 27-NOV-1997.
PF 01-MAY-1997; G01205.
PR 22-MAY-1996; GB-010748.
PR 22-MAY-1996; GB-010687.
PR 22-MAY-1996; GB-010695.
PR 22-MAY-1996; GB-010697.
PR 22-MAY-1996; GB-010738.
PR 22-MAY-1996; GB-010739.
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (ZENE) ZENECA LTD.
PI Christian PD, Claudianos A, Daly A, Duncan RE, Suner M,
PI Windass JD;
DR WPI: 98-018430/02.
PT New nucleic acid encoding a combination of insecticidal sub:unit(s)
PT of wasp toxin - and related transformed cells, insect pathogens and
PT combinations of proteins, useful as insecticides
PS Claim 1; Pages 57-58; 84pp; English.

CC This is a contiguous genomic sequence which encodes part of a spliced RNA
CC that is related to the cDNA comprised by pBrhTX-1(a)1.1. This 1811 base
CC pair spliced RNA derived from the Bracon hebetor genomic clone encodes
CC at least two of the insecticidal toxin subunits shown in sequences
CC W52124-W52128. The spliced RNA can hybridise with extension products
CC prepared from a 564 base pair (V17145) template with 6 specified primers.
CC A nucleic acid encoding at least one of the specified subunits can be
CC modified so that mRNA instability motifs and/or fortuitous splice sites
CC are removed, or insect-pest preference codons are used, so that
CC expression of this nucleic acid in insect cells yields practically the

CC same protein as unmodified nucleic acid in its endogenous organism. The
CC nucleic acid encoding an insecticidal toxin subunit can be complementary
CC to a sequence that hybridises under specified conditions to any of
CC sequences shown in V17145 to V17149. Cells transformed with these
CC nucleic acids, organisms regenerated from these cells and pathogens
CC containing these nucleic acids and insecticidal compositions comprising
CC a combinations of the toxin subunits are all used for control of insects.
CC The nucleic acids are used to produce recombinant baculoviruses for
CC insect control.
SQ Sequence 1811 BP; 649 A; 259 C; 292 G; 611 T;

Query Match 10.8%; Score 34.6; DB 1; Length 1811;
Best Local Similarity 55.4%; Pred. No. 1.9;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 26 TAGTACAGTTACACCAATAGGAACAAGAAAGTTGTGACATGTAGTAGG 85
DB 207 TAAACCTCCTCCCAATACTTAAAAAAGATTCGAAGCTGTTCTAGA 266
OY 86 AGTGTGATCCCTTACTCCCATCAAAAAAATGATACATGCTTAAGATARAAG 145
DB 267 AATGTCCAGTTTGGACGCAAAACGAAGAAACATTGATATAATTGTTAGG 326
OY 146 G 146
DB 327 G 327

RESULT 10
V74675
ID V74675 standard; DNA; 10813 BP.

AC V74675;
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus contig SEQ ID #364.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.

FH Key Location/Qualifiers
FT misc_feature 301..360

FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 2101..2160

FT /tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3901..3960

FT /tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 5701..5760

FT /tag- d
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 7501..7560

FT /tag- e
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 9301..9360
FT /tag- f

FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT EP-786519-A2.
PN 30-JUL-1997.
PD 07-JAN-1997; 100117.
PF 05-JAN-1996; US-009861.
PR (HUMA-) HUMAN GENOME SCI INC.
PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI; 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1248-1254; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 10813 BP; 3513 A; 1895 C; 1594 G; 3451 T;

Query Match 10.7%; Score 34.2; DB 1; Length 10813;
Best Local Similarity 48.6%; Pred. No. 3.6;
Matches 90; Conservative 1; Mismatches 94; Indels 0; Gaps 0;

OY 11 CTTCTGTGTAGAAATAGTACAGTTACACCAATAGGAACAAGAAAGTTGT 70
DB 107 CTACATTTTATGAAAAATTCATTTTATACAAATTCACACTTATATGCACTTCAT 166
OY 71 GACATGTGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAATGATATCATGG 130
DB 167 TACAGTTACTTTTATTGTTGATGCTTACATGTTTCTAAAAAAATTTGTTATCATAA 226
OY 131 TTAAGGATARAAGGCAATATTTATCATATGCTTAAAGAGAAGAGAAATATAC 190
DB 227 TTAACGTTGAATTAAGAAAAATTAACCTGGAGATAAAAATGAAATATAAAGATAC 286
OY 191 TACTT 195
DB 287 TAATT 291

RESULT 11
X20248_04
Continuation (5 of 10) of X20248 from base 400001 (Borrelia burgdorferi polynucleotid
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 10.6%; Score 33.8; DB 1; Length 110000;
Best Local Similarity 53.4%; Pred. No. 7.8;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 42 AATAGCAACAACAAAGAAAGTTGTGACATTGTAGTAGGAGTGTGACCCCTTAC 101
DB 14419 AATAATCTCAACAATAAAAAGAAACTATAGCAATAAAATGCTTCTAATAATCTCCCAA 14478
OY 102 TCCCAATCAAAAAAATGATGATGTTAAAGATARAAGGCAATATTTATCATA 161
DB 14479 ACTTCACACATGATATAAATAATAGTAGAAGCTTATAATCTCCCATAGCCACAT 14538
OY 162 TGTCTTAAAGAG 174
DB 14539 CTTCAAAAATAG 14551

RESULT 15
T72883
ID T72883 standard; cDNA; 7295 BP.
AC T72883;
DT 12-SEP-1997 (first entry)
DE Plasmodium var-1 gene.
KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; Immune response;
OS Plasmodium vivax.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 517..7065
FT /*tag= a
FT /product= var-1
PN WO9640766-A2.
PD 19-DEC-1996.
PE 07-JUN-1996; U09508.
PR 07-JUN-1995; US-487826.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnals C, Miller LH, Peterson DS, SIm KL, Su X;
PI Wellens TE;
DR WPI: 97-052231/05.
DR P-PSDB; W22476.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins
PS Claim 4; Page 67-69; 96pp; English.
CC This sequence represents the var-1 gene of Plasmodium. Var-1 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The

CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 7295 BP; 2866 A; 1067 C; 1455 G; 1907 T;

Query Match 10.4%; Score 33.2; DB 1; Length 7295;
Best Local Similarity 51.4%; Pred. No. 5.9;
Matches 74; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

OY 15 TTGTTAGAAATAGTACAGTTACACCAATAGACACACAAAAAGTGTGACA 74
DB 4885 TGGCTAGAAATATTTTTCAGAGATTATATAGAAATTAAGCTATGTACG 4944
OY 75 TTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATACATGTTAA 134
DB 4945 AAAAAGGAAGATGATGCCAATGTATAAAAGCTGTATAGAAAAATGGTACAGAAAAA 5004
OY 135 AGATARAAGGCCAATATTTATC 158
DB 5005 ACGAAAGAATGCGCAAAAAATAAAC 5028

Search completed: September 28, 1999, 12:30:20
Job time: 5440 sec

[illegible]

```

RESULT 2
US-08-441-139-3
; Sequence 3, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wittup, Dr. Karl D.
; APPLICANT: Roblison, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2367 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 251..2176
US-08-441-139-3

```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|--------------|
| Query Match | 10.8%; | Score 34.4; | DB 3; | Length 2367; |
| Best Local Similarity | 64.1%; | Pred. No. 0.27; | | |
| Matches 50; | Conservative 1; | Mismatches 27; | Indels 0; | Gaps 0 |

```

Oy  112 AAAAAAAAAATGCATCATCGTTAAAGGATARAAGGGCAATATTTATCATATGTTCTAAAA 171
    ||| ||||| | ||| ||||| |||:| ||||| | ||||| |||||
Db  2275 AAAGAAATGTCTGGATTTTGAAAAGGTAACACTATGATATTTTTATTTAAATGTTCTAAAA 2334

```

| | | | |
|----|------|----------------------|------|
| OY | 172 | GAGAAGGAGAGAAAAATA | 189 |
| | | | |
| Db | 2335 | AAAAAAAAAAAAAAAAAAAA | 2352 |

```

1      RESULT      3
2      US-08-720-899-1
3      ; Sequence 1, Application US/08720899.
4      ; Patent No. 5753460
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Bisgaard-Frantzen, Henrik
8      ; APPLICANT: Borchert, Torben Vedel
9      ; APPLICANT: Svendsen, Allan
10     ; APPLICANT: Thellersen, Marianne
11     ; APPLICANT: Van der Zee, Pia
12     ; TITLE OF INVENTION: AMYLASE VARIANTS
13     ; NUMBER OF SEQUENCES: 38
14     ;
15     ; CORRESPONDENCE ADDRESS:
16     ; ADDRESSEE: NO. 57534600 NO. 5753460disk of NO. 5753460th America, Inc.
17     ; STREET: 405 Lexington Avenue, 64th Floor
18     ; CITY: New York
19     ; STATE: New York
20     ; COUNTRY: USA
21     ; ZIP: 10174-6401
22     ;
23     ; COMPUTER READABLE FORM:
24     ; MEDIUM TYPE: Floppy disk
25     ; COMPUTER: IBM PC compatible
26     ; OPERATING SYSTEM: PC-DOS/MS-DOS
27     ; SOFTWARE: Patentin Release #1.0, Version #1.30
28     ;
29     ; CURRENT APPLICATION DATA:
30     ; APPLICATION NUMBER: US/08/720,899
31     ; FILING DATE: 10-OCT-1996
32     ; CLASSIFICATION: 435
33     ;
34     ; PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: US 08/343,804
36     ; FILING DATE: 22-NOV-1994
37     ;
38     ; ATTORNEY/AGENT INFORMATION:
39     ; NAME: Lowney Dr., Karen A.
40     ; REGISTRATION NUMBER: 31,274
41     ; REFERENCE/DOCKET NUMBER: 4054,214-US
42     ;
43     ; TELECOMMUNICATION INFORMATION:
44     ; TELEPHONE: 212-867-0123
45     ; TELEFAX: 212-878-9655
46     ;
47     ; INFORMATION FOR SEQ ID NO: 1:
48     ;
49     ; SEQUENCE CHARACTERISTICS:
50     ; LENGTH: 1920 base pairs
51     ; TYPE: nucleic acid
52     ; STRANDEDNESS: single
53     ; TOPOLOGY: linear
54     ;
55     ; FEATURE:
56     ; NAME/KEY: CDS
57     ; LOCATION: 334..1872
58     ;
59     ; FEATURE:
60     ; NAME/KEY: sig_peptide
61     ; LOCATION: 334..420
62     ;
63     ; FEATURE:
64     ; NAME/KEY: mat_peptide
65     ; LOCATION: 421..1869
66     ;
67     US-08-720-899-1

```

| | | | | |
|--------------------------|--------|----------------|-----------|--------------|
| Query Match | 9.98; | Score 31.8; | DB 3; | Length 1920; |
| Best Local Similarity | 56.88; | Pred. No. 1.4; | | |
| Matches 54; Conservative | 2; | Mismatches 39; | Indels 0; | Gaps 0; |

OY 110 AAAAAAAAAATGCATCATGTTAAAGATARAAGGGCAATTTTATCATATGTTCTAA 169
| | | | | | | | : | | | | | | | |
Db 187 AGAAGCAGAGAGCGCTATTGAATAAATAGTAGAAGCGCCCATATCGGGCGTTTCCTTTGG 246

| | | | | |
|----|-----|-------------------|------------------|-----|
| QY | 170 | AAGAGAGGAAGAGAAAT | TACTACTTTCTCRAAT | 204 |
| | | | | |
| Db | 247 | AAGAAATATAGGAAAT | TGTTACTTGTATAAAT | 281 |


```

RESULT 4
US-08-459-610-1
; Sequence 1, Application US/08459610
; Patent No. 5801043
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Svendsen, Allan
; APPLICANT: Thellersen, Marianne
; APPLICANT: Van der Zee, Pia
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5801043o No. 5801043disk of No. 5801043th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,610
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/343,804
; FILING DATE: 22-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4054.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..1872
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 334..420
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 421..1869
;
US-08-459-610-1

Query Match          9.9%; Score 31.8; DB 3; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches 54; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

QY 110 AAAAAAATGATACATGTTAAAGCATARAAGGCAATATTTATCATATGTTCTAA 169
   1111111111111111111111111111111111111111111111111111111
DB 187 AGAAGCAGAGAGGCTATTGAATATAAGTAGAAGCGCCATATCGGCGCTTTCTTTGG 246
   1111111111111111111111111111111111111111111111111111111

QY 170 AAGAGAAGGAAGAGAAATACTACTTCTCRAAAT 204
   1111111111111111111111111111111111111111111111111111111
DB 247 AAGAAATATAGGAAATGTTACTGTTAAAAAT 281
   1111111111111111111111111111111111111111111111111111111

RESULT 5
US-08-343-804-1
; Sequence 1, Application US/08343804

```

```

: Patent NO. 5830837
: GENERAL INFORMATION:
: APPLICANT: Bisgaard-Frantzen, Henrik
: APPLICANT: Borchert, Torben Vedel
: APPLICANT: Svendsen, Allan
: APPLICANT: Thellersen, Marianne
: APPLICANT: Van der Zee, Pia
: TITLE OF INVENTION: AMYLASE VARIANTS
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 5830837o No. 5830837disk of No. 5830837th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/343,804
: FILING DATE: 22-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lowney Dr., Karen A.
: REGISTRATION NUMBER: 31,274
: REFERENCE/DOCKET NUMBER: 4054.214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 334..1872
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 334..420
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 421..1869
: US-08-343-804-1

Query Match          9.9%; Score 31.8; DB 4; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches 54; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

QY 110 AAAAAAATGATGATACATGCTTAAAGGATARAAGGCAATATTTATCATATGTTCTAA 169
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 AGAAGCAGAGAGGCGTATGCAATGAATGAGTAGAAGCGCCCAATATCGCGCTTTCTTTGG 246

QY 170 AAGAGAAGGAAGAGAATACTACTTCTCRAAT 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AAGAAATATAGGAAATGCTACTTGTTAAAAAT 281

RESULT 6
US-07-867-106-2
: Sequence 2, Application US/07867106
: Patent No. 5389526
: GENERAL INFORMATION:
: APPLICANT: Slade, Martin B
: APPLICANT: Chang, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium

```


NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

```

Query Match          9.9%; Score 31.6; DB 1; Length 5852;
Best Local Similarity 51.5%; Pred. No. 2.4;
Matches 70; Conservative 1; Mismatches 65; Indels 0; Gaps 0;

QY      36  ACAACCATAGGACACACAAGAAAGTGTGTGACATGTAGTAGGAGTGTACC 95
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5449  AAAAAAAAAAAAAAAAAAAAAAAAAATCAATATGTTTATGTGTTTAAGATTTTTAATC 5508

QY      96  CCTTACTCCCCATCAAAAAAAAAATGATACATGGTTAAAGGATARAAGGCAATATTT 155
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5509  TCGTCAATGATTTTAAATAATAAATCGATACATAATTTTAAAAAAAACCCCTTTACATTTT 5568

QY      156  ATCATATGTTCTAAA 171
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5569  TTATTTTAAATTCCAA 5584

RESULT      7
US-08-485-139-1/c
; Sequence 1, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; TITLE OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;

```

```

ADDRESS: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,139
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-096
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4059 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: C1 gene of Zea mays

FEATURE:
NAME/KEY: -
LOCATION: 279..284
OTHER INFORMATION: /label= HpaI

FEATURE:
NAME/KEY: -
LOCATION: 447..452
OTHER INFORMATION: /label= EcoRI

FEATURE:
NAME/KEY: -
LOCATION: 1735..1740
OTHER INFORMATION: /label= AatII

FEATURE:
NAME/KEY: -
LOCATION: 1505..1510
OTHER INFORMATION: /label= EcoRI

FEATURE:
NAME/KEY: -
LOCATION: 2081..2086
OTHER INFORMATION: /label= XhoI

FEATURE:
NAME/KEY: -
LOCATION: 2418..2430
OTHER INFORMATION: /label= SfiI

FEATURE:
NAME/KEY: -
LOCATION: 2669..2674
OTHER INFORMATION: /label= SnaBI

FEATURE:
NAME/KEY: -
LOCATION: 2634..2639
OTHER INFORMATION: /label= SnaBI

FEATURE:
NAME/KEY: -
LOCATION: 3008..3013
OTHER INFORMATION: /label= HpaI

FEATURE:
NAME/KEY: -
LOCATION: 1..1077
OTHER INFORMATION: /label= Pci
OTHER INFORMATION: /note= "region containing promoter of C1 gene"
FEATURE:

```


NAME/KEY: -
LOCATION: 1078..2134
OTHER INFORMATION: /label= C1
OTHER INFORMATION: /note= "coding region of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 2135..2430
OTHER INFORMATION: /label= 3'C1
OTHER INFORMATION: /note= "region containing polyadenylation signal"
OTHER INFORMATION: of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 1033..1038
OTHER INFORMATION: /label= TATA-Box
FEATURE:
NAME/KEY: -
LOCATION: 1061..1062
OTHER INFORMATION: /label= transcript-init
OTHER INFORMATION: /note= "transcription initiation site"
FEATURE:
NAME/KEY: Intron
LOCATION: 1211..1299
FEATURE:
NAME/KEY: Intron
LOCATION: 1430..1575
FEATURE:
NAME/KEY: -
LOCATION: 935..939
OTHER INFORMATION: /label= C1-S
OTHER INFORMATION: /note= "TGCAG sequence (in C1 gene) which in the
OTHER INFORMATION: C1-S sequence is changed to TTAGG"
US-08-485-139-1

Query Match 9.7%; Score 31; DB 4; Length 4059;
Best Local Similarity 61.7%; Pred. No. 3.1;
Matches 66; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

QY 18 TTAGAAATAGTACGTACACCAATAGACAAACAAAAGTTGTGACATG 77
DB 3636 TTATAAATGACACATTTTCATCATATAGAGACAA-AATAAGAGAAAGTTGTGAGATCA 3578

QY 78 TAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGAT 124
DB 3577 AGTCTAAATAAGTCCCATGAGTCAACTTATAAAAAAGATGAT 3531

RESULT 8
US-08-308-872B-5/C
Sequence 5, Application US/08308872B
Patent No. 5661006
GENERAL INFORMATION:
APPLICANT: BROWN, Thomas David Key
APPLICANT: HORSBURGH, Brian Colin
TITLE OF INVENTION: CANINE CORONAVIRUS SUBUNIT VACCINE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5661006 Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,872B
FILING DATE: 19-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,846

FILING DATE: 05-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,641
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91.303.737.0
FILING DATE: 25-APR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4435 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Canine corona virus
STRAIN: CCV-V54
FEATURE:
NAME/KEY: CDS
LOCATION: 60..4418
OTHER INFORMATION: /label= CCV-C54_Spikegene
US-08-308-872B-5

Query Match 9.6%; Score 30.6; DB 2; Length 4435;
Best Local Similarity 47.5%; Pred. No. 4.1;
Matches 87; Conservative 1; Mismatches 95; Indels 0; Gaps 0;

QY 2 CCCTGAGGCTTCTGTAGAAATAGTACAGTTACAACCAATAGAACACAAAAGAA 61

DB 429 CTCTATAGGCTGATATGTAACCAATGATCTACAGGACTACCAATGATCTAGTA 370

QY 62 AAAGTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATG 121

DB 369 AAGTTTACTCTTGATGCTGACAGTACTATTGCCATAGCTTCCATATCAAAATAAATG 310

QY 122 GATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAGAGAGAAAG 181

DB 309 CATGATGTTACTAAATAATGTAAGCTGTAGTTGTTGCTGTTCTGGAACAGTTGTACC 250

QY 182 AGA 184
DB 249 ACA 247

RESULT 9
US-08-520-678A-23/C
Sequence 23, Application US/08520678A
Patent No. 5874565
GENERAL INFORMATION:
APPLICANT: Rice, Charles M.
APPLICANT: Kolykhalov, Alexander A.
TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howell & Haferkamp, L.C.
STREET: 7733 Forsyth Blvd., Suite 1400
CITY: St. Louis
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/520,678A
;; FILING DATE:
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Henderson, Melodie W.
;; REGISTRATION NUMBER: 37,848
;; REFERENCE/DOCKET NUMBER: 6029-6836
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 314-727-5188
;; TELEFAX: 314-727-6092
;;
;; TELEEX:
;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 321 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-520-678A-23
;
Query Match 9.6%; Score 30.6; DB 4; Length 321;
Best Local Similarity 60.0%; Pred. No. 1.6;
Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
;
QY 20 AGAAATAGTACAGTACACCAATAGGACACAAAAAGTTGTGACATTGTA 79
Db 170 AAAAAAAAAAAAAAAAAAAAAAAAAAGGAGAAAAAAAAAAAAAAAAACAGGAATGCCCTAAGA 111
;
QY 80 GTAGGAGTGTGTACCCCTTACTCC 104
Db 110 GGCCGGAGTGTTCACCCCACTTC 86
;
RESULT 10
US-08-323-170B-1
; Sequence 1, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

;; LENGTH: 9636 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 149..9556
;; US-08-323-170B-1
;
Query Match 9.5%; Score 30.4; DB 3; Length 9636;
Best Local Similarity 52.5%; Pred. No. 6.2;
Matches 64; Conservative 1; Mismatches 57; Indels 0; Gaps 0;
;
QY 108 TCATAAAAAAAAAATGCATACATGCTTAAGGATARAAGGCAATATTTCATATGTTCT 167
Db 4379 TCATAAACATGTGGAATAATCAATTAGGAGGAAAAAGTTATATATCATATTACTGTTCA 4438
;
QY 168 AAAAGAGAGGAGAGAAAAATACTACTTTCGAAATGGAAGCCCTTAAGGCTTGA 227
Db 4439 AAAAGAGAGCAGAAAGTAAAGTTGTTTCATTGATATGACACATGCACATATGTTAGT 4498
;
QY 228 TA 229
Db 4499 TA 4500
;
RESULT 11
PCT-US92-06300-1
; Sequence 1, Application PC/TUS9206300
; GENERAL INFORMATION:
; APPLICANT: Hurwitz, David R
; APPLICANT: Nathan, Margaret
; TITLE OF INVENTION: Transgenic Protein Production
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer, Inc.
; STREET: 500 Virginia Ave., Bldg. 3A
; CITY: Ft. Washington
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19034
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06300
; FILING DATE: 19920730
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Rosanne
; REGISTRATION NUMBER: 52,534
; REFERENCE/DOCKET NUMBER: A0856-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 962-4130
; TELEFAX: (215) 962-4107
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19557 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PUBLICATION INFORMATION:
; AUTHORS: Minghetti, P P
; AUTHORS: Rufiner, D E
; AUTHORS: Kuang, W.-J.

Db 10757 ATATATATATATATATATTCATCAATATAAAAAATGAAAACTTTAAAAATATATACA 10816
QY 169 AAAGAGAAGAGAAGAAATACTACTT 195
Db 10817 GAGATGAAATTAACATAAATAGTAATT 10843

RESULT 13

US-08-449-315-36
Sequence 36, Application US/08449315
Patent No. 5650505

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/449,315
FILING DATE: 24-MAY-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/181,271

FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197

FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566

FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/165,667

FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847

FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441

FILING DATE: 21-DEC-1990

* PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/GC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase genomic DNA
IMMEDIATE SOURCE:
CLONE: pbscucchrchts
US-08-449-315-36

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;
Matches 73; Conservative 1; Mismatches 73; Indels 0; Gaps 0;

QY 49 ACAACAAAAAGAAAGTTGTGACATTTGTAGTAGGAGTGTGTACCCCTTACTCCCAT 108
Db 10697 ACAGCAAAAAAAGAAATGCAAGGTTGAGTACATTTGTATAGATGCAATATATAT 10756
QY 109 CAAAAAATGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTA 168
Db 10757 ATATATATATATATATATTCATCAATATAAAAAATGAAAAATCTTTAAATAATATACA 10816
QY 169 AAAGAGAAGAGAAGAAATACTACTT 195
Db 10817 GAGATGAAATTAACATAAATAGTAATT 10843

RESULT 14

US-08-444-803-36
Sequence 36, Application US/08444803
Patent No. 5654414

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice

APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,803
FILING DATE: 19-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase genomic DNA
IMMEDIATE SOURCE:
CLONE: pBscucchrht5
US-08-444-803-36

Query Match 9.38; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;
Matches 73; Conservative 1; Mismatches 73; Indels 0; Gaps 0;

OY 49 ACAACAAAAGAAAAGTTGTGACATGTAGTAGGAGTGTGACCCCTTACTCCCAT 108
DB 10697 ACAGCAAAAAAATGCAAGGTTGAGTACCATGTGCATAGTGCATATATAT 10756
OY 109 CAAAAAATGATGATCATGCTTAAGGATARAAGGCAATATTTATCATATGTTCTA 168
DB 10757 ATATATATATATATATATATTCATCATATATAAAAAATTGAAATCTTAAATAATATATACA 10816
OY 169 AAAGAGAAGAGAAGAAATACTACTT 195
DB 10817 GAGATGAATTAACATAAGTAAT 10843

RESULT 15
US-08-449-043-36
Sequence 36, Application US/08449043
Patent No. 5689044

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
IMMEDIATE SOURCE:
CLONE: pBScucchrcht5
US-08-449-043-36

| | Matches | 73; | Conservative | 1; | Mismatches | 73; | Indels | 0; | Gaps | 0; |
|----|---------|--|--------------|----|------------|-----|--------|----|------|----|
| QY | 49 | ACAACAAAAGAAAGTTGTGCATGTAGAGGAGTGCTACCCCTACTCCCAT | 108 | | | | | | | |
| Db | 10697 | ACAGCAAAAAAATAATGCAAGGCTGAGTACCATGTGCATAGATGCTCAATATATAT | 10756 | | | | | | | |
| QY | 109 | CAAAAAAATGATACATGCTTAAGCATARAAGGCAATATTTCATCATATGCTCTA | 168 | | | | | | | |
| Db | 10757 | ATATATATATATATATATATTCATCAATATAAAAAATGAAAAATCTTTAAATAATATATACA | 10816 | | | | | | | |
| QY | 169 | AAAGAGAAGGAGAGAAATACTACTT | 195 | | | | | | | |
| Db | 10817 | GAGATGAAATTAACTAAATAGTAATT | 10843 | | | | | | | |

Search completed: September 28, 1999, 11:34:23
Job time: 2143 sec

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;

This Page Blank (uspto)

256 TCCTCCTTARAGTGCATGACTTGAC 283
||| :::: |||||:::
121 sProThriLeAlaLysHisAspLeu 130

seq_name: p1r2:A60944

seq_documentation_block:

cytochrome b - Leishmania mexicana amazonensis mitochondrion (SGC6) (fragment)
C/Species: mitochondrion leishmania mexicana amazonensis
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C/Accession: A60944
R/Lee, S.T.; Tarn, C.; Wang, C.Y.
Mol. Biochem. Parasitol. 56, 197-208, 1992
A/Title: Characterization of sequence changes in kinetoplast DNA maxicircles of drug-res
A/Reference number: A60944
A/Accession: A60944
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-255 <LEE>
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SGC6
C/Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocuinol
C/Keywords: mitochondrion
F;10-255/Domain: cytochrome b homology (fragment) <CBH>
F;10-198/Domain: cytochrome b6 homology <CB6>
F;211-255/Domain: plastocuinol--plastocyanin reductase 17K protein homology (fragment) <

alignment_scores:
Quality: 64.00 Length: 86
Ratio: 1.455 Gaps: 6
Percent Similarity: 51.163 Percent Identity: 30.233

alignment_block:

US-09-030-606-224/rev x A60944 ..

Align seg 1/1 to: A60944 from: 1 to: 255

289 TACCGTGTCCAGTCAATGCACTTTAAAGGAGATGACGGCCACATTG 240
|||::: :::: :: ||||| ::|||
61 TyrgluileuileValIglyValIleArgArg....SerHisIleCy 75
239 TGTCTTCAGTATCAAGCACCCTTAAAGGCTTCATTGAGAAAGTAG 190
|::: ::||| ||| ::|||
75 sleuHisIstYrtyrPhePhePheMetLeuHis.....I 88
189 TATTTCTCTCTCTCTCTCT....TTAGACATATGATAAATATGTC 146
::|||::||| ||::: |||
88 IephelysleuHisSerIleAsnAsnIleTrpHisThrtyrPheValCys 104
145 CCTTYTATCCTTTAACCATGATATCCATTTTCTTTGATGGGAGTAGG 96
||| :::: |||::: |||::: |||
105 ...Gly.ArgPheCysAsnIleTyrlleTyrserArgIleGlyPheIleG 120
95 GGTAACACACTCCCTACTACATATGCACAACTTTTCTTTGTGTTCC 46
|||||::||| |||||
120 lYTyValleuProCysThrMet.....MetSer 129
45 TATTGG 40
|||||
130 TyrTrp 131

seq_name: p1r2:C71697

seq_documentation_block:

hypothetical protein RP396 - Rickettsia prowazekii
C/Species: Rickettsia prowazekii
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 21-Nov-1998
C/Accession: C71697
R/Andersson, S.G.E.; Zomrodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, T.
Nature 396, 133-140, 1998

A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A/Reference number: A71630

A/Accession: C71697
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-208 <AND>
A/Cross-references: GB:AJ235271; GB:AJ235269; NID:g3860788; PID:e1342697; PID:g386095
A/Experimental source: strain Madrid E
C/Genetics:
A/Gene: RP396

alignment_scores:
Quality: 63.50 Length: 66
Ratio: 1.716 Gaps: 2
Percent Similarity: 56.061 Percent Identity: 28.788

alignment_block:

US-09-030-606-224 x C71697 ..

Align seg 1/1 to: C71697 from: 1 to: 208

83 GGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGAT..... 124
|||::: ||||| |||::: |||
93 GlyGlyLysTyrrProTyrgluProleuLysThrArgValAspLysSerAr 109
125 ...ACATGGTTAAAGATARAAGGCGCATATTATCATATGTTCTAAA 170
||| ::|||::: |||
109 gPheThrAsnLysSerAspSerSerPheThrGlySerGlyIleAlaThrT 126
171 AGAGAGGAAGAGAAATACTACTTTCTCRAAATGGAAGCCCTTAAAGT 220
::: ||| ::||| |||::: |||
126 hrAsnAsnLysGlyGluTyrrtyrPheIleSerMetLeuProTyrrLysSer 142
221 GCTTGATACTGAAGGACACAAATGTGGCGTC...CATCCTCTTTA 265
|||::: |||::: |||::: |||
143 ProArgTyrrLeuArgAsnAlaAsnIleArgIleGluHisProAsnLeu 158

seq_name: p1r2:I48134

seq_documentation_block:

ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - American spiny rat (Ec
C/Species: mitochondrion Echimyis chrysurus (American spiny rat)
C/Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 24-Jul-1998
C/Accession: I48134
R/da Silva, M.N.; Patton, J.L.
Mol. Phylogenet. Evol. 2, 243-255, 1993
A/Title: Amazonian phylogeography: mtDNA sequence variation in arboreal echmyid rode
A/Reference number: A49605; MUID:94184505
A/Accession: I48134
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-379 <RES>
A/Cross-references: GB:I23341; NID:g995848; PID:g995849
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SGC1
C/Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocui
C/Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative ph
F;11-339/Domain: cytochrome b homology <CBH>
F;11-209/Domain: cytochrome b6 homology <CB6>
F;221-339/Domain: plastocuinol--plastocyanin reductase 17K protein homology <17K>

alignment_scores:
Quality: 63.50 Length: 83
Ratio: 1.628 Gaps: 4
Percent Similarity: 46.988 Percent Identity: 26.506

alignment_block:

US-09-030-606-224/rev x I48134 ..

Align seg 1/1 to: I48134 from: 1 to: 379


```
217 TTTAAGGCTTCATTTYGAGAAAGTAGTATTTCTCTCTCTCTTT 168
    ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 PhepNealaphenHis.....PheValLeuProPheileil 189
167 AGACATATGATAAATATATGCCCTTATCTTAAACATGTATCCATT 118
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
189 eThralaMetVal.....MetileHisl 197
117 TTTTGTGATGGGAGTAGGGGTACACACT..... 86
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 euleuPheleuHlsgluThrglySerAsnAsnProSerGlyLeuAsnSer 213
85 .....CCCTACTAC.....AATGTCAC 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 AspSerAspLysIleProPheHlsProTyrTyrThrIleLysAspIlele 230
68 AAACCTTTCTTTTGTGTCCTATTTGGTTGTAACGTACTATTCT 20
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 uGlyLeuMetPheMetLeuPheAlaLeuMetMetLeuLeuPheSer 246
```

seq_name: p1r2:I48180

seq_documentation_block:

ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - American spiny rat (Isoth
C:Species: mitochondrion Isothrix bistriatus (American spiny rat)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 24-Jul-1998
C:Accession: I48180
R:da Silva, M.N.; Patton, J.L.
Mol. Phylogenet. Evol. 2, 243-255, 1993
A:Title: Amazonian phylogeography: mtDNA sequence variation in arboreal echimyid rodents
A:Reference number: A49605; MUID:94184505
A:Accession: I48180
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-379 <RES>
A:Cross-references: GB:L23355; NID:g995846; PID:g995847
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastoquinol
C:Keywords: chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative phosph
F:I1-209/Domain: cytochrome b6 homology <CBH>
F:I1-209/Domain: cytochrome b6 homology <CBH>
F:221-339/Domain: plastoquinol--plastocyanin reductase 17K protein homology <17K>

alignment_scores:

| Quality: | 63.50 | Length: | 83 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.587 | Gaps: | 4 |
| Percent Similarity: | 48.193 | Percent Identity: | 27.711 |

alignment_block:

US-09-030-606-224/rev x I48180 ..

Align seg 1/1 to: I48180 from: 1 to: 379

```
217 TTTAAGGCTTCATTTYGAGAAAGTAGTATTTCTCTCTCTCTTT 168
    ||| ::||| ||| ||| ||| ||| ||| ||| ||| |||
178 PhepNealaphenHis.....PheileuProPheileil 189
167 AGACATATGATAAATATATGCCCTTATCTTAAACATGTATCCATT 118
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
189 eThralaMetVal.....MetileHisl 197
117 TTTTGTGATGGGAGTAGGGGTACACACT..... 86
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 euleuPheleuHlsgluThrglySerAsnAsnProSerGlyLeuAsnSer 213
85 .....CCCTACTACATGTCACAAAC..... 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 AspSerAspLysIleProPheHlsProTyrTyrThrIleLysAspIlele 230
```

```
64 ....TTTTCTTTTGTGTCCTATTTGGTTGTAACGTACTATTCT 20
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 uGlyLeuMetPheMetLeuLeuLeuLeuLeuThrleuValLeuPheSer 246
```

seq_name: p1r1:A44217

seq_documentation_block:

genome polyprotein - bovine viral diarrhoea virus (strain SD-1)
C:Species: bovine viral diarrhoea virus, BVDV
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-May-1998
C:Accession: A44217
R:Deng, R.; Brock, K.V.
Virology 191, 867-879, 1992
A:Title: Molecular cloning and nucleotide sequence of a pestivirus genome, noncytopat
A:Reference number: A44217
A:Accession: A44217
A:Molecule type: genomic RNA
A:Residues: 1-3898 <DEN>
A:Cross-references: GB:M96751; NID:g289507; PID:g289508
A>Note: this polyprotein may be cleaved into several mature proteins, including p20 p
and p75 protein; the cleavage sites are not reported
C:Superfamily: pestivirus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; glycoprotein; P-loop; polyprotein; RNA binding; zinc finger
F:2-234/Product: viral proteinase p20 #status predicted <VPT>
F:253-265/Region: hydrophobic
F:546-1115/Product: major envelope glycoprotein gp55 #status predicted <EGP>
F:547-562/Region: hydrophobic
F:556-670/Region: hydrophobic
F:675-694/Region: hydrophobic
F:1031-1046/Region: hydrophobic
F:1074-1099/Region: hydrophobic
F:1149-1164/Region: hydrophobic
F:1217-1238/Region: hydrophobic
F:1252-1269/Region: hydrophobic
F:1271-1292/Region: hydrophobic
F:1293-1304/Region: hydrophobic
F:1357-1373/Region: hydrophobic
F:1484-1512/Region: zinc finger
F:1815-2098/Domain: DEAD/H box helicase homology <DEAD>
F:1815-1822/Region: nucleotide-binding motif A (P-loop)
F:1906-1911/Region: nucleotide-binding motif B
F:1910-1913/Region: DEAD motif
F:2562-2582/Region: hydrophobic
F:272,281,296,335,365,370,413,487,597,809,878,922,990,1357,1419,1713,2134,2217,2494,2

alignment_scores:

| Quality: | 63.50 | Length: | 125 |
|---------------------|--------|-------------------|--------|
| Ratio: | 0.992 | Gaps: | 7 |
| Percent Similarity: | 51.200 | Percent Identity: | 24.000 |

alignment_block:

US-09-030-606-224/rev x A44217 ..

Align seg 1/1 to: A44217 from: 1 to: 3898

```
320 GTCACATGCKGAGTYTAACAGTCAACAGTTACCGTGTCACAGTCATGC 271
    ||| ::||| ||| ||| ||| ||| ||| ||| ||| |||
958 ValilealaThrAspThrLysleuGlyProMetProCysLysProHlsgl 974
270 AACCTTAAGGAGGATGGA.....CGGCCACATTTGTGCTCTCAG 230
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 uileleSerSerGlyGlyProileGluLysThr.AlaCysThrPheAsn 990
229 TATCAAAAGCCTTTAAGGCTTCATTTYGAGAAAGTAGTATTTCTCT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
991 TyrThrArgThrLeuLysAsnLysTyrPheGlu..... 1001
179 TCCTTCT.....CTTTAGACATATGATAAAA.....TATGCG 145
    ||| ::||| ||| ||| ||| ||| ||| ||| ||| |||
1002 .ProArgAspSerTyrPheGlnGlnTyrMetLeuLysGlyAspTyr.... 1016
144 CTTYTATCCCTTAACCATGTATCCATTTTGTGATGGGAGTAGAGG 95
```



```

1017 .....GlnTyrTyrPheaspLeuGluValThr 1025
      |||
94  GTACACACTCCCTACTACAAATGTCACAACTTTTC..... 59
      |||
1026 ASPHISHISARGASPTyrPhealacIleuSerileuValvalValal 1042
      |||
58 .....TTTGTGTTCCTATTGGTGTGAACGTACTAT 25
      |||
1042 aLeuLeuGlyGlyArgTyrValleuTyrPheuleuValThrTyrMetVal 1059
      |||
24  TTTCTAACACAGAGCCCTCAGGG 2
      |||
1059 euSerGluGlnIlysalasSerGly 1066

```

seq_name: p1r2:A45554

```

seq_documentation_block:
  rhoptry complex polypeptide RhopH3 - Plasmodium falciparum
C/Species: Plasmodium falciparum
C/Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Feb-1995
C/Accession: A45554
R/Brown, H.J.; Coppel, R.L.
Mol. Biochem. Parasitol. 49, 99-110, 1991
A/Title: Primary structure of a Plasmodium falciparum rhoptry antigen.
A/Reference number: A45554; MUID:92131054
A/Contents: rhoptry
A/Accession: A45554
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-895 <BRO>
A/Note: sequence extracted from NCBI backbone (NCBIN:77766, NCBIIP:77768)

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 63.00 | Length: 84 |
| Ratio: | 1.400 | Gaps: 3 |
| Percent Similarity: | 53.571 | Percent Identity: 26.190 |

alignment block:
US-09-030-606-224/rev x A45554 ..

Align seg 1/1 to: A45554 from: 1 to: 895

```

247 CACATTTGTGTCCTTCAGTATCAAAAGCCTTTAAGGGCTTCATTTGGA 198
    |||:: |||::: |||::: |||:: |||:: |||:: |||:: |||::
533 HisValLeuValValLysProHisTyrThrPheHisProPheTyrThrG1 549
197 GAAAGTAGTATTTCTCTTCCTTCCTT.....TTAGAAC 163
    |::: |::: |::: |::: |::: |::: |::: |:::
549 uGluIleTyrIlePheLeuLysHisLeuIleIleIleLeuG1LeuSerA 566
162 ATATGATAAATATATTGCCCTTAACTCTTAACCATGTATCCATTTT 113
    ::::: ::::: |||::: |||
566 snLeuValSerSerAsnProSer.....SerProPhe 576
112 TTTCATGGGGAGTAAGGGGTACACACTCCCTACTACATGTC..ACAA 66
    ||||| |::: |||||
577 PheAspThrIleIleGluPheLeuValThrTyrTyrAsnLysGlySerG1 593
65 CTTTTCCTTTTGTGTGTCCTATGTTGTTGTAAGTGTACTATTTCTACA 16
    : ||| ||| |||::: ::::: |||||::: |||
593 uLysPheValLeuTyrPheIleSerIleIleSerValLeuTyrIleAsnG 610
15 AG 14
::
610 lu 610

```

seq_name: plr2:T00379

```
seq_documentation_block:
KIAA0640 protein - human (fragment)
C:Species: Homo sapiens (man)
```

C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 01-Feb-1999
C/Accession: T00379
R/Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998
A/Title: Prediction of the coding sequences of unidentified human genes. X. The compl
A/Reference number: Z14142
A/Accession: T00379
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-603 <1st>
A/Cross-references: EMBL:AB014540; NID:d1204295; PID:d1032576
C/Genetics:
A/Gene: KIAA0640

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 62.50 | Length: 63 |
| Ratio: | 2.155 | Gaps: 3 |
| Percent Similarity: | 46.032 | Percent Identity: 34.921 |

alignment_block: . .
US-09-030-606-224 x T00379 .

Align seg 1/1 to: T00379 from: 1 to: 603

```

115 AAAATGGATACATGGTTAAAGCATARAAGGC..... 147
    |||:::|||||||:::||| |||||
231 LysGlnGlyTyrMetMetLys.....LysGlnHisArgArgLysAsnTr 245
148 .....AATATTGTATCATATGTT... 165
    |||||:::|||||
245 pThrGluArgTyrPheValLeuLysProAsnIleIleSerTyrTyrValS 262
166 .....CTAAAGAGAAGAGAAGAAATACTACTTTCRRAATGCA 207
    |||||:::||||: |||||::: |||
262 erGluAspLeuLysAspLysLysGlyAspIleLeuLeuAspGluAsnCyS 278
208 AGCCCTTAAGGTGCTTTGATACTGTGAAGCACACAATGT 246
    ||| ::||| |||||
279 CysValGluSerLeuProAspLysAspGlyLysLysCyS 291

```

seq_name: p1r1:NBMSH

```

seq_documentation_block:
complement factor H precursor - mouse
N:Alternate names: protein beta-1-H
C:Species: Mus musculus (house mouse)
C:date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 20-Mar-1998
C:Accession: A26154; I49711; I49728
R:Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length
A:Reference number: A26154; MUID:86233353
A:Accession: A26154
A:Molecule type: mRNA
A:Residues: 1-1234 <KRI>
A:Cross-references: GB:M12660; NID:g193724; PID:g387181
R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the complement
A:Reference number: I49711; MUID:90111033
A:Accession: I49711
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-18 <RES>
A:Cross-references: GB:M31979; NID:g193726; PID:g193729
R:Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.
Biochemistry 28, 9891-9897, 1989
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
A:Reference number: I49728; MUID:90148935
A:Accession: I49728
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

```


A;Residues: 1-19 <RE2>

A;Cross-references: GB:J02891; NID:g193805; PID:g553926

C;Comment: Two codominant alleles of factor H are present in mice.

C;Comment: Factor H functions as a cofactor in the inactivation of C3b by serine proteinase (C5 convertase) in the alternative complement pathway.

C;Genetics:

A;Map position: 1

C;Superfamily: complement factor H; complement factor H repeat homology

C;Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-1234/Product: complement factor H #status predicted <MP1>

F;21-80/Domain: complement factor H repeat homology <FH01>

F;85-141/Domain: complement factor H repeat homology <FH02>

F;146-205/Domain: complement factor H repeat homology <FH03>

F;210-262/Domain: complement factor H repeat homology <FH04>

F;246-248/Region: cell attachment (R-G-D) motif

F;267-320/Domain: complement factor H repeat homology <FH05>

F;325-385/Domain: complement factor H repeat homology <FH06>

F;389-442/Domain: complement factor H repeat homology <FH07>

F;448-505/Domain: complement factor H repeat homology <FH08>

F;509-564/Domain: complement factor H repeat homology <FH09>

F;569-622/Domain: complement factor H repeat homology <FH10>

F;629-683/Domain: complement factor H repeat homology <FH11>

F;690-743/Domain: complement factor H repeat homology <FH12>

F;752-802/Domain: complement factor H repeat homology <FH13>

F;808-861/Domain: complement factor H repeat homology <FH14>

F;867-931/Domain: complement factor H repeat homology <FH15>

F;936-989/Domain: complement factor H repeat homology <FH16>

F;994-1048/Domain: complement factor H repeat homology <FH17>

F;1053-1107/Domain: complement factor H repeat homology <FH18>

F;1114-1168/Domain: complement factor H repeat homology <FH19>

F;1172-1233/Domain: complement factor H repeat homology <FH20>

F;21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

alignment_scores:

Quality: 62.00 Length: 87

Ratio: 1.409 Gaps: 8

Percent Similarity: 50.575 Percent Identity: 29.885

alignment_block:

US-09-030-606-224 x NBMSH ..

Align seg 1/1 to: NBMSH from: 1 to: 1234

96 CCTTACTCCCA.....TCAAAAAATATGAT 124

||||: ||| ::|||:||||:|

1057 ProHisValProAsnAlaThrIleValThrArgThrLysAsnLysTyrLe 1073

125 ACATGGTTAAGGATA.....RAAG 144

||||||| |||:|

1073 uHisGlyAspArgValArgTyrGluCysAsnLysProLeuGluLeupheg 1090

145 GCGAATATTTCATATGTTCTAAAGAGAGAGAGAAATACTACT 194

||||| ::|||:|

1090 LysGlnValGluValMetCys.....GluAsnGlyIle 1100

195 TTCTCRAAATGGAAGCCCTTAAGGTGCTTGATGACTGAACACAAAT 244

||||| ||| ||| |||:|

1101 TrpThrGlu...LysPro...LysCysArgAspSerThrGly...LysC 1114

245 GTGGCGGTCATCTCTTARAGTTGCATGACTTGACACGTAACGT 294

||||| ||| |||:|

1114 YsGlyProPro.ProProIleAspAsnGlyAspIleThrSerLeuSerie 1130

295 TGCAGTT 301

: |||

1130 uProVal 1132

seq_name: p1r2:A70122

seq_documentation_block:

Glucose inhibited division protein B (gldb) homolog - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998

C;Accession: A70122

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh

son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

; Bowman, C.; Garland, S.; Fujit, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943

A;Accession: A70122

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-208 <KLE>

A;Cross-references: GB:AE001128; GB:AE000783; NID:g2688057; PID:g2688060; TIGR:BB0177

A;Experimental source: strain B31

alignment_scores:

Quality: 61.50 Length: 91

Ratio: 1.255 Gaps: 4

Percent Similarity: 53.846 Percent Identity: 25.275

alignment_block:

US-09-030-606-224 x A70122 ..

Align seg 1/1 to: A70122 from: 1 to: 208

47 GAACACAAAGAAAGAAAGTTGTGACATGTAGTAGGAGGTGTACCC 96

||| ::|:| |||:|

68 GluIleAsnProSerGluValLeuAspValGlySerGlyAlaGlyPhePr 84

97 C.....TTACTCC 104

| |||||

84 OGlyIleIleLeuAlaIlePheAspSerSerArgLysTyrTrpLeuLeuG 101

105 CCATCAAAAAAATGATACATGTTAAG..... 136

::||| |||:|

101 LuArgSerLysLysSerThrPheLeuLysMetIleLysLeuGluLeu 117

137CATARAAGGCAATATTTCATATGTTCTTAAAGAGAG 177

::|:| |||:| |||:|

118 AspLeuGluAsnValLysIleLeuGluTyrGluIle.....GluLysG 132

178 GAAGAGAAATACTACTTCTCRAAATGAAGCCCTTAAGGT..... 220

||||:||||| |||:|

132 uLysLysLysTyrGluPheIleThrIleArgAlaPheArgAsnMetAsnG 149

221GCTTGATCTGAGAGAC 238

||||| |||:|

149 LuTyrAlaLeuIleLeuLysAsn 156

seq_name: p1r2:B39218

seq_documentation_block:

nicotinic acetylcholine receptor alpha-5 chain precursor - chicken

C;Species: Gallus gallus (chicken)

C;Date: 30-Aug-1991 #sequence_revision 30-Aug-1991 #text_change 17-Mar-1999

C;Accession: B39218; S68498

R;Couturier, S.; Erkman, L.; Valera, S.; Rungger, D.; Bertrand, S.; Boulter, J.; Ball

J. Biol. Chem. 265, 17560-17567, 1990

A;Title: alpha5, alpha3, and non-alpha3. Three clustered avian genes encoding neurona

A;Reference number: A39218; MUID:91009210

A;Accession: B39218

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-454 <COU>

A;Cross-references: GB:J05643

R;Ramirez-Latorre, J.; Yu, C.R.; Qu, X.; Perin, F.; Karlin, A.; Role, L.

Nature 380, 347-351, 1996

A;Title: Functional contributions of alpha-5 subunit to neuronal acetylcholine recept

A:Reference number: S68498; MUID:96176246
A:Accession: S68498
A:Molecule type: mRNA
A:Residues: 264-288 <RAM>
C:Superfamily: acetylcholine receptor
C:Keywords: glycoprotein; ion channel; neurotransmitter receptor; transmembrane protein
F:237-260/Domain: transmembrane #status predicted <TM1>
F:268-286/Domain: transmembrane #status predicted <TM2>
F:302-323/Domain: transmembrane #status predicted <TM3>
F:416-434/Domain: transmembrane #status predicted <TM4>
F:156-170,220-221/Disulfide bonds: #status predicted
F:169/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
Quality: 61.50 Length: 102
Ratio: 1.183 Gaps: 4
Percent Similarity: 50.980 Percent Identity: 29.412

alignment_block:

US-09-030-606-224/rev x B39218 ..

Align seg 1/1 to: B39218 from: 1 to: 454

```
315 AATGCKGAGTCTTAAACTGCAACAGTTACCGTGTCACAGTCATGCACTY 266
      |||::||| :::: |||::||| |||::||| ::::
201 AsnGlyIuTrpGluIleValThrAlaThrGlySerIyGlyAsnArgTh 217
      : |||
265 TAAAGGAGGATGGACGGCCACATTGTGCTCCTCAGTATC.....A 225
      : |||
217 rAspGlyCysCysTrpTyrProPheValThrTyrSerPheIleIleArgA 234
      : |||
224 AAGCACCTTAAAGGGCTTCCATTGAGAGAAAGTAGTATTTCTCTCCTT 175
      :: |||
234 rGluPProLeuPhe.....TyrThrLeuPheLeu 243
      : |||
174 CTCTTTAGACATATGATAAAATATGCCCCCTTATCCTTAACCATGT 125
      : |||
244 IleIleProCysIle.....GlyLeuSerPheLeuThrVa 255
      : |||
124 ATCCATTTTCTTGTATGGGAGTAAGGGGTACACACTCCCTACTACAA 75
      : |||
255 lLeuValPheTyr.....LeuProSerAsnG 264
      : |||
74 TGTCAACAACTTTTCTTTTGTGTTCTTATGTTGTAACGTACTAT 25
      : |||
264 lUGlyGluIySileSerLeuCythrSerValLeuValSerLeuThrVal 280
      : |||
24 TTTCTA 19
      |||
281 PheLeu 282
```

seq_name: plr2:T03053

seq_documentation_block:
ribonucleoside-diphosphate reductase large chain homolog - Chilo Iridescent virus
C:Species: Chilo Iridescent virus
C:Date: 24-Mar-1999 #sequence_revlsion 24-Mar-1999 #text_change 24-Mar-1999
C:Accession: T03053
R:Bahr, U.; Tidona, C.A.; Darai, G.
Virus Genes 15, 235-245, 1997
A:Title: The DNA sequence of Chilo Iridescent virus between the genome coordinates 0.101
A:Reference number: Z14834
A:Accession: T03053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-959 <BAH>
A:Cross-references: EMBL:AF003534; NID:g2738385; PID:g2738400

alignment_scores:
Quality: 61.00 Length: 75
Ratio: 1.694 Gaps: 4

Percent Similarity: 48.000 Percent Identity: 30.667

alignment_block:

US-09-030-606-224/rev x T03053 ..

Align seg 1/1 to: T03053 from: 1 to: 959

```
295 AACAGTTACCGTGTCGAAGTCATGCACTYTAAGAGAGATGACGGCCA 246
      |||:: |||::|||::|||::|||::|||
358 AsnAlaGluArgValAspAlaGlnAspLeuLysProArgMet..... 371
      : |||
245 CATTTGTGCTCTCAGTATCAAGCACCTTTAAGGCTTCCATTGAGAGA 196
      : |||
372 .....LysL 373
      : |||
195 AAGTAGTATTTTCTCTCTCTCTTTA.....GAACATATGATA 155
      : |||
373 euIleLysPheAspLeuProThrLeuPheGlyAsnSerGluHisAspIle 389
      : |||
154 AAATATGCCCCCTTATATCCTTAACCATGTATCCATTTTCTTGATGG 105
      : |||
390 LysTyr.....ProTyrThrHisGlyPhePheCysG 400
      : |||
104 GGAGTAAGGGGTACACACTCCCTAC 80
      : |||
400 yAsp...GlyThrTyrThrLysTyr 407
```

OM of: US-09-030-606-224 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_SPOOL/US09030606/runat_24091999_171618_29883/app_query.fasta.1
-DB=SwissProt_37 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdl -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-224

Query length: 320

Database: SwissProt_37:*

Database sequences: 77977

Database length: 28268293

Search time (sec): 155.430000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|----------------------------|-----------|--------|--------|------|----------------------------------|
| SwissProt_37:IF3_MYCGE + | 71.00 | 152.16 | 0.3518 | 184 | P47438 mycoplasma genitalium. |
| SwissProt_37:SECS_YEAST - | 66.50 | 127.34 | 1.61 | 971 | P89102 saccharomyces cerevis |
| SwissProt_37:PARE_STAUA + | 66.00 | 129.45 | 1.80 | 663 | P50072 staphylococcus aureus |
| SwissProt_37:IF3_MYCPN + | 65.00 | 137.37 | 2.15 | 201 | P78024 mycoplasma pneumoniae |
| SwissProt_37:CYB_CHEMA - | 63.50 | 128.41 | 3.59 | 379 | P034169 chelirogaleus major (16 |
| SwissProt_37:POLG_BYDVS - | 63.50 | 108.38 | 4.56 | 3898 | P001499 bovine viral diarrhoe |
| SwissProt_37:ACH5_CHICK - | 62.50 | 124.52 | 4.94 | 454 | P26152 gallus gallus (chicken) |
| SwissProt_37:GIDB_BORBU + | 61.50 | 128.90 | 6.15 | 208 | P53363 borrelia burgdorferi |
| SwissProt_37:CFAH_MOUSE + | 62.00 | 114.76 | 6.35 | 1234 | P06909 mus musculus (mouse) |
| SwissProt_37:ZEAD_MAIZE - | 60.50 | 125.58 | 8.40 | 233 | P24450 zea mays (maize). ze |
| SwissProt_37:YEC5_YEAST - | 61.00 | 112.75 | 8.54 | 1188 | P39991 saccharomyces cerevis |
| SwissProt_37:DD31_DICDI - | 60.00 | 123.18 | 9.90 | 269 | P002465 dictyostelium discoid |
| SwissProt_37:GYRB_MYCCA + | 60.00 | 115.69 | 10.83 | 643 | P50028 mycoplasma capricolum |
| SwissProt_37:CD14_RABIT - | 59.00 | 118.05 | 13.82 | 372 | P028680 oryctolagus cuniculus |
| SwissProt_37:CYB_DASGE - | 59.00 | 117.85 | 13.85 | 381 | P02604 dasyurus geoffroii. cy |
| SwissProt_37:PURI_HUMAN + | 59.00 | 115.22 | 14.29 | 517 | P006203 homo sapiens (human). |
| SwissProt_37:SYG_METJA + | 59.00 | 114.28 | 14.45 | 577 | P05681 methanococcus jannasch |
| SwissProt_37:CR72_BACTR - | 59.00 | 113.48 | 14.59 | 633 | P21254 bacillus thuringiensis |
| SwissProt_37:ZEAC_MAIZE - | 58.50 | 120.83 | 15.32 | 235 | P24449 zea mays (maize). ze |
| SwissProt_37:VBR1_DICDI + | 58.50 | 120.10 | 15.45 | 256 | P14970 cassava latent virus (|
| SwissProt_37:IF31_DICDI - | 59.00 | 107.16 | 15.73 | 1321 | P015818 dictyostelium discoid |
| SwissProt_37:CYB_HYSAF - | 58.50 | 116.61 | 16.11 | 384 | P04910 hystrix africaeaustral |
| SwissProt_37:P2C2_YEAST - | 58.50 | 114.99 | 16.42 | 464 | P39966 saccharomyces cerevis |
| SwissProt_37:KOM8_YEAST - | 58.00 | 115.65 | 18.67 | 375 | P012222 saccharomyces cerevis |
| SwissProt_37:CYB_GALCU - | 58.00 | 115.51 | 18.70 | 381 | P34868 galeocerdo cuvier (tig |
| SwissProt_37:CYB_NEGBR - | 58.00 | 115.51 | 18.70 | 381 | P34872 negaprior brevirostris |
| SwissProt_37:Y874_HAEIN + | 58.00 | 115.11 | 18.78 | 399 | P44067 haemophilus influenza |
| SwissProt_37:NU6M_CAEEL - | 57.50 | 122.70 | 19.66 | 144 | P24885 caenorhabditis elegans |
| SwissProt_37:Y556_METJA + | 57.50 | 120.55 | 20.17 | 185 | P05796 methanococcus jannasch |
| SwissProt_37:VIS1_HUMAN + | 57.50 | 120.32 | 20.23 | 190 | P42423 homo sapiens (human). |
| SwissProt_37:NTIM_LOCOMI - | 57.50 | 116.03 | 21.29 | 313 | P09093 locusta migratoria (m |
| SwissProt_37:CFH1_HUMAN + | 57.50 | 115.58 | 21.40 | 330 | P003591 homo sapiens (human). |
| SwissProt_37:CYB_NINT1 - | 57.50 | 114.34 | 21.72 | 381 | P020435 ningauli timeleylei. cyt |
| SwissProt_37:CYB_SMITA - | 57.50 | 114.70 | 21.72 | 381 | P020435 smitopsis macrooura |
| SwissProt_37:FUCO_HUMAN - | 57.50 | 112.70 | 22.15 | 461 | P04066 homo sapiens (human). |
| SwissProt_37:GLGB_BACSV - | 57.50 | 110.06 | 22.86 | 627 | P39118 bacillus subtilis. 1, |
| SwissProt_37:CFAH_HUMAN + | 57.50 | 104.26 | 24.49 | 1231 | P08603 homo sapiens (human) |
| SwissProt_37:CYB_DASSP - | 57.00 | 113.17 | 25.23 | 381 | P034382 dasyurus spartacus (na |
| SwissProt_37:AF9_HUMAN - | 57.00 | 109.74 | 26.29 | 568 | P42568 homo sapiens (human). |
| SwissProt_37:PLZF_HUMAN - | 57.00 | 108.28 | 26.75 | 673 | P00516 homo sapiens (human). |
| SwissProt_37:SYV_LACCA + | 57.00 | 105.77 | 27.56 | 901 | P36420 lactobacillus casei. v |
| SwissProt_37:IF3_MYCGE + | 56.50 | 117.04 | 27.61 | 212 | P005426 mycoplasma fermentans. |

| | | | | | |
|---------------------------|-------|--------|-------|-----|--------------------------|
| SwissProt_37:ZEAF_MAIZE - | 56.50 | 115.97 | 27.96 | 240 | P04703 zea mays (maize). |
| SwissProt_37:ZEAF_MAIZE - | 56.50 | 115.97 | 27.96 | 240 | P06676 zea mays (maize). |
| SwissProt_37:ZEAF_MAIZE - | 56.50 | 115.97 | 27.96 | 240 | P06677 zea mays (maize). |

seq_name: SwissProt_37:IF3_MYCGE

seq_documentation_block:

| ID | IF3_MYCGE | STANDARD | PRT | 184 AA. |
|----|--|----------|-----|---------|
| AC | P47438 | | | |
| DT | 01-FEB-1996 (REL. 33, CREATED) | | | |
| DT | 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE) | | | |
| DT | 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE) | | | |
| DE | TRANSLATION INITIATION FACTOR IF-3. | | | |
| GN | INFC OR MG196. | | | |
| OS | MYCOPLASMA GENITALIUM. | | | |
| OC | BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES; | | | |
| OC | MYCOPLASMATACEAE; MYCOPLASMA. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-ATCC 33530 / G-37; | | | |
| RX | MEDLINE; 96026346. | | | |
| RA | FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A., | | | |
| RA | FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M., | | | |
| RA | FRITCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L., | | | |
| RA | NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M., | | | |
| RA | TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S., | | | |
| RA | PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.; | | | |
| RT | "The minimal gene complement of Mycoplasma genitalium."; | | | |
| RL | SCIENCE 270:397-403(1995). | | | |
| CC | -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE | | | |
| CC | EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN | | | |
| CC | FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S | | | |
| CC | SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS. | | | |
| CC | -1- SUBUNIT: MONOMER (BY SIMILARITY). | | | |
| CC | -1- SUBCELLULAR LOCATION: CYTOPLASMIC. | | | |
| CC | -1- SIMILARITY: BELONGS TO THE IF-3 FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; U39697; G1045881; ALT_INIT. | | | |
| DR | TIGR; MG196; | | | |
| DR | PROSITE; P500938; IF3; 1. | | | |
| DR | PFAM; PF00707; IF3; 1. | | | |
| DR | HSSP; P03000; 1TIF. | | | |
| KW | INITIATION FACTOR; PROTEIN BIOSYNTHESIS. | | | |
| SQ | SEQUENCE 184 AA; 21276 MW; 6584A264 CRC32; | | | |

alignment_scores:

| Quality: | 71.00 | Length: | 76 |
|---------------------|--------|-------------------|--------|
| Ratio: | 1.821 | Gaps: | 3 |
| Percent Similarity: | 51.316 | Percent Identity: | 31.579 |

alignment_block:

US-09-030-606-224 x IF3_MYCGE ..

Align seg 1/1 to: IF3_MYCGE from: 1 to: 184

```
56 AAAGAAAAGTTGTGACATTGATAGGAGTGTGATACCCCTTACTCC 105
:::|||||  |||:::||||:::  |||  |||
52 GlnGluysGlnLeuAspLeuValLeuIleAlaProAsnProThrLysPr 68
106 CATCAAAAAAATGATACATGTTAAAGATARAAGGCAATATTT 155
|||  |||  |||  |||
68 olleVallys.....LeuLeuAspPheGlyArgTyrThrT 80
156 ATCATATGTTCTAAAGAGAGAAGAGAATAACTACTTCTCRAAATG 205
```


SQ SEQUENCE 663 AA; 74364 MW; E3682180 CRC32;

alignment_scores:

Quality: 66.00 Length: 72
Ratio: 1.650 Gaps: 3
Percent Similarity: 55.556 Percent Identity: 30.556

alignment_block:

US-09-030-606-224 x PARE_STAUV ..

Align seg 1/1 to: PARE_STAUV from: 1 to: 663

```

43 ATAGGACACACAAAAGAAAGTTGTGACATGTAGTAGGAGTGTGT 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
29 IleglyserThrasplysArgGlyLeuHSHSLeuValTyrGluIleVa 45
93 ACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGGATARA 142
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
45 lAspAsnSerValAspGluValLeuAsnGlyTyr..... 56
143 AGGCAATATTTATCATATGTCTTAAAGAGAGAGAGAAATACTA 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 .GlyAsnGluIleAspValThrIle..AsnLysAspGlySerIleSer 71
193 CTTTCTCRAAAATGAGAGCCCTTAAAGTGCTTGATACTGAAGACACA 242
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
72 lIleGluAspAsnGly.....ArgGlyMetProThrGlyIleHisLy 85
243 ATGTGGCCGTCATCC 258
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
85 sSerGlyLysProThr 90

```

seq_name: SwissProt_37:IF3_MYCPN

seq_documentation_block:

```

ID IF3_MYCPN STANDARD; PRT; 201 AA.
AC P78024;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRANSLATION INITIATION FACTOR IF-3.
GN INFC.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT *Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.*;
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE
CC EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN
CC FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S
CC SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE IF-3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE000005; G1673688; -.
CC PROSITE; PS00938; IF3; 1.

```

DR PEAM; PF00707; IF3; 1.
DR HSSP; P03000; 1TIF.
KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS.
SQ SEQUENCE 201 AA; 23145 MW; 1FC1283C CRC32;

alignment_scores:

Quality: 65.00 Length: 76
Ratio: 1.667 Gaps: 3
Percent Similarity: 51.316 Percent Identity: 30.263

alignment_block:

US-09-030-606-224 x IF3_MYCPN ..

Align seg 1/1 to: IF3_MYCPN from: 1 to: 201

```

56 AAGAAAGTTTGTGACATGTAGTAGGAGTGTACCCCTACTCCC 105
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
69 GluGluLysGlnLeuAspLeuValLeuIleGlySerAsnProAlaLysPr 85
106 CATCAAAAAAATGATACATGCTTAAAGATARAAGGCCAATATT 155
||| | || | || | || | || | || | || | || | || | || |
85 oIleValLys.....LeuLeuAspPheGlyArgTyrThr 97
156 ATCATATGTTCTTAAAGAGAGAGAGAAATACTACTTCTCRAAATG 205
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
97 yAspLeu.....LysArgLysLysArg..... 104
206 GAAGCCCTTAAAGTGCTTGTGACTGAAGACACAAATGTGCCGTCA 255
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
105 GlnSerLysLysAsnGlnThrIleIleGlnIleLysGluValValVally 121
256 TCCTCCTTARAGTTGCATGACTTGAC 283
|| |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 sProThrIleAlaLysHisAspLeuGlu 130

```

seq_name: SwissProt_37:CYB_CHEMA

seq_documentation_block:

ID CYB_CHEMA STANDARD; PRT; 379 AA.

AC Q34169;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CYTOCHROME B (EC 1.10.2.2).

GN MTCYB OR COB OR CYTB.

OS CHEIROGALEUS MAJOR (LEMUR).

OC MITOCHONDRION.

OC EUKARYOTA; METAZOA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; STREPSIRHINI; CHEIROGALEIDAE; CHEIROGALEUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96209867.

RA YODER A.D., CARTMILL M., RUVOLO M., SMITH R., VILGALYS R.;

RT *Ancient single origin for Malagasy primates.*;

RL PROC. NATL. ACAD. SCI. U.S.A. 93:5122-5126(1996).

CC -1- FUNCTION: COMPONENT III OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.

CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +

CC 2 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.

CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,

CC CYTOCHROME C1 AND THE RIESKE PROTEIN.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

alignment_scores:

Quality: 60.50 Length: 66
Ratio: 1.681 Gaps: 1
Percent Similarity: 54.545 Percent Identity: 30.303

alignment_block:

US-09-030-606-224/rev x ZEAD_MAIZE ..

Align seg 1/1 to: ZEAD_MAIZE from: 1 to: 233

```
209 CTTCCATTGAGAGAGATATTCTCTCTCTCTTTAGACATA 160
|||||::: ||| ::||| |||
169 LeuProPheSerGlnLeuAlaAspValSerProAlaAlaPheLeuThrG1 185
:::|||||::: |||
159 TGATAAATATGCCCCCTTATCCTTTAACCATGTATCCATTTT 110
:::|||||::: |||
185 nGlnGlnLeuProPheTyrLeuHisAlaMet..... 196
109 GATGGGAGTAAGGGGTACACACTCCCTACTACATGTACAAACTTTT 60
:::|||||::: |||
197 .....ProAsnIaGlyThrLeuLeuGlnLeuGlnLeuLeu 209
59 CTTTGTGTCTATGTTGTACTGTACTATTTCTACACAGAA 12
||| |||::: ||| |||:::
210 ProPheAsnGlnLeuAlaLeuThrAsnSerThrValPheTyrGlnGln 225
```

seq_name: SwissProt_37:YEC5_YEAST

seq_documentation_block:

ID YEC5_YEAST STANDARD; PRT; 1188 AA.

AC P39991;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 136.1 KD PROTEIN IN YMA3-RIP1 INTERGENIC REGION.
GN YEL025C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOEGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U18530; G602392; -
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 878 898 POTENTIAL.
FT TRANSMEM 1089 1109 POTENTIAL.
SQ SEQUENCE 1188 AA; 136068 MW; 80AD1745 CRC32;

alignment_scores:

Quality: 61.00 Length: 63
Ratio: 1.649 Gaps: 3
Percent Similarity: 58.730 Percent Identity: 33.333

alignment_block:

US-09-030-606-224/rev x YEC5_YEAST ..

Align seg 1/1 to: YEC5_YEAST from: 1 to: 1188

```
253 GACGGCCACATTTGTGCTTCAGTATCAAGACCTTTAAGGCTTCCA 204
|||||::: ||| ::||| |||
729 AspSerArgIleSerIlePheGlnPheGlySerThrPheTyrArgGluG1 745
203 TTTGAGAAAGATATTCTCTCTCTCTCTTTAGACATATGATAA 154
||| ||| ::| ||||| |||||::: |||
745 yPheThrLysGlnMetThrSerLeuLysProGlnLeuGlnAsnHisIleP 762
153 AATATTGCCCTTATCCTTTAACCATGTATCCATTTTTTTGAATGG 104
||| |||::: ||| |||::: |||
762 heTyrLeuProSerIle.....Pro.ValGluLeu..... 771
103 GAGTAAGGGGTACACACTCCCTACTACATGTACAA 67
|||||::: |||::: |||
772 .....TyrThrValProAsnAsnLeuAsnGln 780
```

seq_name: SwissProt_37:DD31_DICDI

seq_documentation_block:

ID DD31_DICDI STANDARD; PRT; 269 AA.

AC Q02465;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE PRESPORE PROTEIN DD31.
GN SPIA.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RX MEDLINE; 91184531.
RA RICHARDSON D.L., HONG C.B., LOOMIS W.F.;
RT "A prespore gene, Dd31, expressed during culmination of Dictyostelium
RT discoideum.";
RL DEV. BIOL. 144:269-280(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE; 92275349.
RA RICHARDSON D.L., LOOMIS W.F.;
RT "Disruption of the sporulation-specific gene spia in Dictyostelium
RT discoideum leads to spore instability.";
RL GENES DEV. 6:1058-1070(1992).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE INNER FACE OF SPORE
CC COAT.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY PRIOR TO SPORE
CC FORMATION AND IN SPORES DURING CULMINATION.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X54452; G7164; -
DR PIR: S16671; S16671.
DR DICTYDB: DD04007; SPIA.
SQ SEQUENCE 269 AA; 30620 MW; 55B3C183 CRC32;

alignment_scores:

Quality: 60.00 Length: 109
Ratio: 1.154 Gaps: 7
Percent Similarity: 47.706 Percent Identity: 29.358

alignment_block:

US-09-030-606-224/rev x DD31_DICDI ..

Align seg 1/1 to: D031_DICDI from: 1 to: 269

```
290 TTACCGTGTCAGATCATGCACACTYTAAGAGAGATGAGCGGCACATTT 241
:::|||||::: ||| ::::: ::: |||
72 ValProAlaProAlaThrAlaHisProArgGlnAlaAsnProSerSe 88
240 GTGTCCTTCAGTATCAAGACACCTTTAAGGCTTCATTTYGAGAAAGTA 191
||| ::| ||| ::| |||
88 rTyrPrometIleArgLeuAsnLeu.....GlulGlu 99
190 GTATTTCTCTCCCTCTCTCTTTAGACATATGATAAATATGCGCCTTY 141
::| ||| ::| ||| ::| |||
99 sple.....AlaIleArgGlnTyrArgGlnIleValLysPhe 111
140 TATCCTTAACCATGTATCCATTTTGTGATGGGAGTAAG..... 96
||| ||| ||| ::|
112 .....GlyIlePheValPheLeuTyrGlnAlaAlaAlaLe 123
95 .....GGTACACACTCCCTACTACATGTC 71
||| ||| ||| |||
123 uValTyrAsnTyrPvalValSerIleGlyThr.IleValTyrSerAlaVal 139
70 ACAACCTTTTCTTTTGTGTGTCCTATGCTGTA.....ACTGT 30
||||| ||| ||| |||
140 AspAsnPhenPheLeuAlaLeuPheTyrMetIleValGlyValProThrIe 156
29 ACTATTTCTAACAAGACCTT 7
: ||||| ||||| |||||
156 u.TyrPheLeuThrArgLysLeu 163
```

seq_name: SwissProt_37:GYRB_MYCCA

seq_documentation_block:
ID GYRB_MYCCA STANDARD; PRT; 643 AA.

```
AC P50028;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE DNA GYRASE SUBUNIT B (EC 5.99.1.3).
GN GYRB.
OS MYCOPLASMA CAPRICOLUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC CAPRICOLUM GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 27343;
RX MEDLINE; 95129856.
RA SANO K.-I., MIYATA M.;
RT "The gyrb gene lies opposite from the replication origin on the
RT circular chromosome of Mycoplasma capricolum.";
RL GENE 151:181-183(1994).
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDANT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28808; G938029; .
DR EMBL; D21231; G533332; .
```

```
DR EMBL; D26016; G532280; .
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
DR PFAM; PF00204; DNA_topoisom; 1.
DR PFAM; PF00986; DNA_gyraseB_C; 1.
DR HSSP; P06982; 1AJ6.
KM TOPOISOMERASE; ISOMERASE; ATP-BINDING.
SQ SEQUENCE 643 AA; 72333 MW; 093FD91D CRC32;
```

alignment_scores:
Quality: 60.00 Length: 72
Ratio: 1.622 Gaps: 2
Percent Similarity: 51.389 Percent Identity: 30.556

alignment_block:
US-09-030-606-224 x GYRB_MYCCA ..

Align seg 1/1 to: GYRB_MYCCA from: 1 to: 643

```
43 ATAGAACACAAAGAAAGAGTTGTGACATTGTAGTAGGAGTGTGT 92
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
29 IleGlySerThrAspAsnLysGlyLeuHisHisLeuValTyrGlnIleVal 45
93 ACCCCTTACTCCCATCAAAAAAATGATACATGCTTAAGATARA 142
|::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|
45 LaspAsnAlaIleAspGlnAlaLeuAlaGlyTyrCysThrGln..... 59
143 AGGCAATATTTTATCATATGCTTTAAAGAGAGAGAGAGAAATACTA 192
|||::: |||||::|::|::|::|::|::|::|::|::|::|::|::|
60 .....IleAspValIleLeuGlnLysAspAsnSerIleThr 71
193 CTTTCTCRAATGGAAGCCCTTAAGGTCTTGATACTGAAGACACAA 242
::: ||||| |||::: ||| |||||
72 ValIleAspAsnGly.....ArgGlyIleProThrGlyMetHisLys 85
243 ATGTGCGCGTCATCC 258
| |||::|::|::|::|::|::|::|::|::|::|::|::|::|
85 sthrgLysProThr 90
```

seq_name: SwissProt_37:CD14_RABIT

seq_documentation_block:
ID CD14_RABIT STANDARD; PRT; 372 AA.

```
AC Q28680;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MONOCYTE DIFFERENTIATION ANTIGEN CD14 PRECURSOR.
GN CD14.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92268819.
RA LEE J.D., KATO K., TOBIAS P.S., KIRKLAND T.N., ULEVITCH R.J.;
RT "Transfection of CD14 into 70z/3 cells dramatically enhances the
RT sensitivity to complexes of lipopolysaccharide (LPS) and LPS binding
RT protein.";
RL J. EXP. MED. 175:1697-1705(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-MYELOID;
RA ISHIDA T., SETOGUCHI M., MATSUURA K., YASUNORI H., AKIZUKI S.,
RA YAMAMOTO S.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: SERVES AS AN LPS RECEPTOR CONTROLLING CELL ACTIVATION
CC UNDER PHYSIOLOGICAL CONDITIONS. WHEN LPS BINDS TO CD14 THE CELLS
CC BECOME ACTIVATED AND RELEASE CYTOKINES AND UP-REGULATE CELL
CC SURFACE MOLECULES, INCLUDING ADHESION MOLECULES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M85233; G164865; -
DR EMBL; D16545; D1022626; -
DR PFAM; PF00560; LRR; 3.
KW ANTIGEN; GLYCOPROTEIN; GPI-ANCHOR; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 372 CD14 DIFFERENTIATION ANTIGEN.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 152 152 POTENTIAL.
FT CARBOHYD 279 279 POTENTIAL.
FT CARBOHYD 320 320 POTENTIAL.
SQ SEQUENCE 372 AA; 39992 MW; CECDDFA8 CRC32;

alignment_scores:
Quality: 59.00 Length: 86
Ratio: 1.283 Gaps: 4
Percent Similarity: 53.488 Percent Identity: 30.233

alignment_block:
US-09-030-606-224/rev x CD14_RABIT ..

Align seg 1/1 to: CD14_RABIT from: 1 to: 372

302 AAAGTGCACAGTACCGGTCCAGTCAT.....GCAACTYT 265
|||||:||||| ||| |||:||||| :|||
305 LysLeuAsnAlaProGlnProGluLeuProLysValValAsnLe 321
264 AAAGGAGATGACGGCCACATTGTGTCCTTCAGATCAAGCACCTTT 215
|::| |||||:|||| ||| |||:||||| |||
321 uSerLeuAspGlyAsnProPheLeuValProGlyAlaSerLysLeuGln 338
214 AAGGCTTCATTYAGAGAAGTAGATTTCTCTCTCTCTCTTTAGA 165
||| :|||:|||||:||||| :|||
338 LysAspLeu.....ThrAsnSerGlyValPheProAla..... 348
164 ACATATGATAAATATGCGCTTATCCTTTACCATGATCATTTT 115
||| ||| |||||:|||||
349CysProProSerProLeuAlaMet..... 356
114 TTTTGTATGGGAGTAGGGGTACACACTCCCTACTACATGTCAACAAC 65
|||:| ||||| :||| |||||
357GlyMetSerGlyThrLeuAlaLeuGlnGlyAlaArgG 370
64 TTTTCTCT 57
|||:|
370 LysPheIle 372

seq_name: SwissProt_37:CYB_DASGE

seq_documentation_block:

ID CYB_DASGE STANDARD; PRT; 381 AA.
AC 020604;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CYTOCHROME B (EC 1.10.2.2).
GN MTCYB OR COB OR CYTB.
OS DASYURUS GEOFFROII.
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYUROMORPHIA; DASYURIDAE; DASYURUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KRAJEWSKI C., YOUNG J., BUCKLEY L., WOOLLEY P.A., WESTERMAN M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
RT cytochrome b, 12S rRNA, and protamine P1 gene trees.";

RL J. MAMMAL. EVOL. 4:217-236(1997).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: O₂H(2) + 2 FERRICYTOCHROME C = O₂ +
CC 2 FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF010266; G2352861; -
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
DR PFAM; PF00032; cytochrome_b_C; 1.
DR PFAM; PF00033; cytochrome_b_N; 1.
KW MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW HEME.
FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 381 AA; 42902 MW; C37AABF4 CRC32;

alignment_scores:
Quality: 59.00 Length: 75
Ratio: 1.439 Gaps: 5
Percent Similarity: 54.667 Percent Identity: 30.667

alignment_block:
US-09-030-606-224/rev x CYB_DASGE ..

Align seg 1/1 to: CYB_DASGE from: 1 to: 381

185 TTCCTCTCCTCTCTTTAGACATATGATAAATATGCCCTTAT.. 138
||| :||| ||| |||||
178 PhePheAlaPheHisPhe.....IleLeuProPheIleI 189
137 ...CCTTACCATGATCCATTTTGTGATGGGAGTAAGGGCTA 92
|||||:||||| :||| |||
189 eThrProLeuAlaIle.ValHisLeuLeuPheLeuHisGluThrGlySer 205
91 CACACT.....CC 84
:|||||
206 AsnAsnProSerGlyIleAsnProAspSerAspLysIleProPheHisPr 222
83 CTACTACATGTCAACAAC.....TTTCTTTTGTGTTCTAT 43
|||||:||||| :||| |||||:|||||
222 OTYTYTThrIleLysAspAlaLeuGlyPheMetLeuLeuLeuIleI 239
42 TGGTGTAACTGTACTATTTCT 20
|||:|
239 euleuLeuAlaLeuPheSer 246

This Page Blank (uspto)

| | | | |
|----|---|---------|---------------------------|
| OC | Potamotrygon. | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | MARTIN A.P., LOVEJOY N., BERMINGHAM E.; | | |
| RT | "Marine Incursion into South America."; | | |
| RL | Nature 396:421-422(1998). | | |
| DR | EMBL; AF110626; AAD12291.1; - | | |
| KW | Mitochondrion. | | |
| FT | NON_TER | 1 | |
| FT | NON_TER | 255 | 255 |
| SO | SEQUENCE | 255 AA; | 29047 MW; E8F7E237 CRC32; |

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 71.50 | Length: 84 |
| Ratio: | 1.589 | Gaps: 4 |
| Percent Similarity: | 53.571 | Percent Identity: 30.952 |

alignment_block:
US-09-030-606-224/rev x 099907 ..

Align seg 1/1 to: 099907 from: 1 to: 255

```

188 ATTTCTCTCTCTCTCT.....TTTAGAAC 163
    |||:::|||||  ||| ||
121 ILETpGLyGLyPhneSeValAspAsnAlaThrLeuThrArgpPheheth 137
    |||:::|||||  ||| ||
162 ATATGATAAATAATGCCCTTAT.....CCTTAACCATGTATCCAT 119
    ||::: |||  ||| ||| ||| |||
137 rPhehIsrPheLeupPheProPheMetIleAlaAlaLeuThrmEt.IlehIs 153
    ||::: |||  ||| ||| ||| |||
118 TTTT TTTTGATGGGAGTAAGGGGTACACACTCC..... 83
    ::::|||  ||| ||| ::::|||
154 LeuLeupPheLeuHIsGluAlaGlySerAsnAsnProThrGlyLeuThrSe 170
    ::::|||  ||| ||| ::::|||
82 .....TACTACAATGTCA 70
    ||| :::::
170 rAspMetAspLysIleGlnPhehIsrProTyrTyrTrpTyrLysAspLeuV 187
    ||| :::::
69 CAAACSTTTTCTTTTGTGTCTCTATATGGTGTGAACGTACTATTTCT 20
    ::::||| ||| ||| ||| ::::||| ||| ::::|||
187 alGlyPhePhePheLeuLeuLeuLeuAlaIleLeuAlaLeuPheThr 203

```

```
seq_documentation_block:
ID 099906 PRELIMINARY; PRT; 210 AA.
```

```

DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBrel. 10, last sequence update)
DT 01-MAY-1999 (TREMBrel. 10, last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Potamotrygon orbigny1.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Rajiformes; Myliobatoidel; Potamotrygonidae;
OC Potamotrygon.
RN [1]
RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BIRMINGHAM E.;
RT "Marine Incursion Into South America.";
RL Nature 396:421-422(1998).
DR EMBL, AF110625; AAD12290.1; -.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 210 210
SQ SEQUENCE 210 AA; 23926 MW; 7235786B CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 70.50 | Length: 83 |
| Ratio: | 1.602 | Gaps: 4 |
| Percent Similarity: | 53.012 | Percent Identity: 31.325 |

```

alignment_block:
US-09-030-606-224/rev x 099906      ..
Align seg 1/1   to: 099906   from: 1   to: 210

```

```

188 ATTTCTCTCTCTCTCT.....TTTAGAAC 163
      |||:::|||||
121 ILETPGLGLYPHESERVALASPASNALATHRLEUTHARGPHEPETH 137
      |||:::|||||
162 ATATGATAAATATATGCCCTTTAT.....CCTTAACCATGTATCCAT 119
      |||:::|||||
137 rPHEHISPhELEUPHepROPHemETILEAlAlAlAlEUTHrMET. ILEHIS 153
      |||:::|||||
118 TTTTNTTTTGATGGGAGTAGAGGGTACACACTCCC..... 83
      |||:::|||||
154 LEUeUPHleuHISglAlAlAGlySERASnSPROTHrGLyEUTHrSE 170
      |||:::|||||
82 .....TACTACAATGTCA 70
      |||:::
170 rASpMetASpLYSILeGLINPhENISPROTYRtyrTYrTYrLYSAsPLEU 187
      |||:::
69 CAACACTTTTCTTTTGTGTGTCTATTTGTTGTGTAAGTGTACTATT 23
      |||:::|||||
187 ALGyPHePhePheLeuLeuLeuLeuLeuAlAlLEuLeUTHrLeUPHe 202
      |||:::|||||

```

seq_name: sp_organelle:Q34293

```

seq_documentation_block:
ID      034293      PRELIMINARY;      PRT;      244 AA.
AC      Q34293;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE      CYTOCHROME B (FRAGMENT).
OS      Dasyatis akajei.
OG      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC      Elasmobranchii; Rajiformes; Myliobatoidel; Dasyatidae; Dasyatis.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      KITAMURA T., TAKEMURA A., WATABE S., TANIUCHI T., SHIMIZU M.;
RT      "Molecular phylogeny of the sharks and rays.";
RL      Fisheries Sci. 0:0-0(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      KITAMURA T., TAKEMURA A., WATABE S., TANIUCHI T., SHIMIZU M.;
RT      "Molecular phylogeny of the sharks and rays of superorder squala
RT      based on mitochondrial cytochrome b gene.";
RL      Fisheries Sci. 62:340-343(1996).
DR      EMBL; D50027; BAA08754.1; -.
DR      PFAM; PF00032; cytochrome_b_C; 1.
DR      PFAM; PF00033; cytochrome_b_N; 1.
KW      Mitochondrion.
FT      NON_TER      1
FT      NON_TER      244
SO      SEQUENCE      244 AA; 27814 MW; EBBB32CE CRC32;

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 70.50 | Length: 102 |
| Ratio: | 1.282 | Gaps: 5 |
| Percent Similarity: | 53.922 | Percent Identity: 28.431 |

```
alignment_block:
US-09-030-606-224/rev x Q34293 ...
```

Align seg 1/1 to: Q34293 from: 1 to: 244

```

230 GATATCAAGCACCCCTTTAAGGCGCTTCATTTYGAGAAAGTAGT..... 189
      :::::  |||  ::::|:::
12 1lethrasnleuSerAlaIeuproTYrileglyaspthrleuValgl 28

```



```
188 ...ATTTCCTCTCTCTCT...TTTA 167
    |||:|||||
28 ntrpIetrgIyGlyPheSerIleaspAsnalatThrLeuthrArgPheP 45
    |||:|||||
166 GAACATATGATAAATATGCCCCCTTAT...CCTTAACCATGTAT 123
    |||:|||||
45 hetHrPheHISpHeuPheProPheValIleAlaLeuThrMet.11 61
    |||:|||||
122 CCATTTTTCATGAGGAGTAAAGGGTACACACTCC... 83
    |||:|||||
61 eHISLeuLeuPheUHSISgluThrglySerAsnAsnProThrglyLeuT 78
    |||:|||||
82 .....TACTACAAT 74
    |||:|||||
78 hrSerAsnThrAspLysIleProPheHISProTyrPheThrTyrLysasp 94
    |||:|||||
73 GTCACAACCTTTTCTTTTGTGTGTCCTATGCTGTGTAACGTACTATT 24
    |||:|||||
95 leuValGlyPhePheIleLeuLeuPheLeuThrLeuLeuAlaLeuPh 111
    |||:|||||
23 TTCT 20
    |||:|||||
111 eThr 112
```

seq_name: sp_organelle:099908

```
seq_documentation_block:
ID 099908 PRELIMINARY; PRT; 252 AA.
AC 099908;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS PotamoTrygon castexi.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Rajiformes; Myliobatoidel; PotamoTrygonidae;
OC PotamoTrygon.
RN [1]
RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.;
RT "Marine incursion into South America.";
RL Nature 396:421-422(1998).
DR EMBL; AF110627; AAD12292.1; -.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 28712 MW; 353314BE CRC32;
```

```
alignment_scores:
Quality: 70.50 Length: 83
Ratio: 1.602 Gaps: 4
Percent Similarity: 53.012 Percent Identity: 31.325
```

alignment_block:

US-09-030-606-224/rev x 099908 ..

Align seg 1/1 to: 099908 from: 1 to: 252

```
188 ATTTCTCTCTCTCTCT...TTTAGAAC 163
    |||:|||||
118 IletrgIyGlyPheSerValaspAsnalatThrLeuthrArgPhePheTh 134
    |||:|||||
162 ATATGATAAATATGCCCCCTTAT...CCTTAACCATGTATCAT 119
    |||:|||||
134 rPheHISpHeuPheProPheMetIleAlaLeuThrMet.11eHIS 150
    |||:|||||
118 TTTTTCATGAGGAGTAAAGGGTACACACTCC... 83
    |||:|||||
151 leuLeuPheUHSISgluAlaGlySerAsnAsnProThrglyLeuThrSe 167
    |||:|||||
82 .....TACTACAATGTCA 70
```

```
167 rAspMetAspLysIleGlnPheHISProTyrTyrTrpTyrLysAspLeuV 184
    |||:|||||
69 CAACCTTTTCTTTTGTGTGTCCTATGCTGTGTAACGTACTATT 23
    |||:|||||
184 aIGlyPhePheLeuLeuLeuLeuAlaIleLeuAlaLeuPhe 199
    |||:|||||
```

seq_name: sp_organelle:099919

```
seq_documentation_block:
ID 099919 PRELIMINARY; PRT; 255 AA.
AC 099919;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Himantura pacifica.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Rajiformes; Myliobatoidel; Dasysatidae; Himantura.
RN [1]
RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.;
RT "Marine incursion into South America.";
RL Nature 396:421-422(1998).
DR EMBL; AF110638; AAD12303.1; -.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28765 MW; 46159853 CRC32;
```

```
alignment_scores:
Quality: 70.50 Length: 102
Ratio: 1.306 Gaps: 5
Percent Similarity: 52.941 Percent Identity: 27.451
```

alignment_block:

US-09-030-606-224/rev x 099919 ..

Align seg 1/1 to: 099919 from: 1 to: 255

```
230 GTATCAAGCACCTTTAAGGGCTTCATTTGAGAAAGTAGT... 189
    |||:|||||
103 IletHrAsnLeuLeuSerAlaLeuProTyrIleGlyAspThrLeuValG1 119
    |||:|||||
188 ...ATTTCCTCTCTCTCTCT...TTTA 167
    |||:|||||
119 n**ValTrpGlyGlyPheSerValaspAsnalatThrLeuthrGlnPheP 136
    |||:|||||
166 GAACATATGATAAATATGCCCCCTT...TATCCTTAACCATGTAT 123
    |||:|||||
136 hetHrPheHISpHeuPheProPheValIleValAlaLeuThrMet.11 152
    |||:|||||
122 CCATTTTTCATGAGGAGTAAAGGGTACACACTCC... 83
    |||:|||||
152 eHISLeuLeuPheUHSISgluAlaGlySerAsnAsnProThrglyLeuT 169
    |||:|||||
82 .....TACTACAAT 74
    |||:|||||
169 leSerAsnThrAspLysIleGlnPheHISProTyrTyrSerTyrLysglu 185
    |||:|||||
73 GTCACAACCTTTTCTTTTGTGTGTCCTATGCTGTGTAACGTACTATT 24
    |||:|||||
186 leuLeuGlyPhePhePheLeuLeuLeuLeuIleLeuAlaLeuPh 202
    |||:|||||
23 TTCT 20
    |||:|||||
202 eThr 203
```

seq_name: sp_organelle:099920

seq_documentation_block:

| | | | | |
|-----|---|-----------------|-------------------------|-----------------|
| ID | 099920 | PRELIMINARY; | PRT; | 255 AA. |
| AC | 099920; | | | |
| DT | 01-MAY-1999 | (TREMBLrel. 10, | Created) | |
| DT | 01-MAY-1999 | (TREMBLrel. 10, | Last sequence update) | |
| DT | 01-MAY-1999 | (TREMBLrel. 10, | Last annotation update) | |
| DE | CYTOCHROME B | (FRAGMENT). | | |
| OS | Himantura schmardae. | | | |
| OG | Mitochondrion. | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; | | | |
| OC | Elasmobranchii; Rajiformes; Myliobatoidel; Dasysatidae; Himantura. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | MARTIN A.P., LOVEJOY N., BERMINGHAM E.; | | | |
| RT | "Marine Incursion into South America."; | | | |
| RL | Nature 396:421-422(1998). | | | |
| DR | EMBL: AF110639; AAD12304.1; -. | | | |
| KW | Mitochondrion. | | | |
| FT | NON_TER | 1 | 1 | |
| FT | NON_TER | 255 | 255 | |
| SEQ | SEQUENCE | 255 AA; | 28904 MW; | 453BAFF4 CRC32; |

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 70.50 | Length: 84 |
| Ratio: | 1.602 | Gaps: 4 |
| Percent Similarity: | 52.381 | Percent Identity: 30.952 |

alignment_block:
US-09-030-606-224/rev x 099920 ..

Align seg 1/1 to: 099920 from: 1 to: 255

```

188 ATTTCTCTCTCCTCTCT...TTAGAAC 163
    |||:::|||||  ||  ||
121 ILeTrpGlyGlyPheSerValAspAsnAlaThrLeuThrArgPhePheThr 137
    |||:::|||||  |||
162 ATATGATAAATATTTGCCCTT...TATCCTTTAACCATGTATCAT 119
    ||::  :::::|||  |||||  |||||
137 rPheHisPheLeuPheProPheValIleValAlaLeuThrmel.IleHis 153
    ||::  :::::|||  |||||  |||||
118 TTTTTTTTGTATGGGAGTAAGGGGTACACACTCC... 83
    :::::|||  |||  ||  :::::|||
154 LeuLeuPheLeuHisGlyAlaGlySerAsnAsnProThrGlyLeuIleSe 170
    :::::|||  |||  ||  :::::|||
82 .....TACTACAATGTCA 70
    .....  |||  :::::
170 rAsnThrAspPlySILeGlnPheHisProTyrTyrSerTyrTyrLysGluLeuL 187
    .....  |||  :::::
69 CAACATTTTCTTTTGTGTGTCCTATGCTGTGTAAGTACTATATTTCT 20
    ::|||  |||  |||  |||  |||  |||  ::|||  |||  :::
187 euGlyPhePhePheLeuLeuLeuLeuLeuIleLeuLeuAlaLeuPheThr 203
    ::|||  |||  |||  |||  |||  |||  ::|||  |||  :::

```

seq_name: sp_organelle:099904

seq_documentation_block:

PRELIMINARY; PRT; 255 AA.

DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Dasyatis hawaiiensis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes
OC Elasmobranchii; Rajiformes; Myliobatoidel; Dasyatidae; Dasyatis.
RN [1]
RP SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BIRMINGHAM E.;
RT "Marine incursion into South America.";
RL Nature 396:421-422(1998).
DR EMBL; AF110623; AAD12288.1; -.
KW Mitochondrion.
FT *NON_TER* 1

| | | | |
|----|----------|---------|---------------------------|
| FT | NON_TER | 255 | 255 |
| SQ | SEQUENCE | 255 AA; | 28769 MN; 2A435F7B CRC32; |

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 69.50 | Length: 84 |
| Ratio: | 1.544 | Gaps: 4 |
| Percent Similarity: | 53.571 | Percent Identity: 30.952 |

alignment_block:
US-09-030-606-224/rev x 099904 ..

Align seg 1/1 to: 099904 from: 1 to: 255

```

188 ATTTCTCTCTCCCTCTCT.....TTTAGAAC 163
    |||:::|||||
121 ILeTPeIGlyPheSerIleAspAsnAlaThrLeuThrArgPhePheThr 137
    |||:::|||||
162 ATATGATAAATATATGGCCCTTAT.....CCTTAACCATGATCCAT 119
    |||:::|||||
137 rPheHisPheLeuPheProPheValIleAlaAlaLeuThrMet.IleHis 153
    |||:::|||||
118 TTTTTTTGGATGGGAGTAAGGGGTACACACTCC..... 83
    |||:::|||||
154 LeuLeuPheLeuHisGluThrGlySerAsnAsnProThrGlyLeuThrSe 170
    |||:::|||||
82 .....TACTACATGTCA 70
    |||:::|||||
170 rAsnMetAspLysIleProPheHisProTyrPheThrTyrLysAspLeuV 187
    |||:::|||||
69 CAAACTTTTCTTTGTGTGTTCTATTGGTTGTAACGTACTATTCT 20
    |||:::|||||
187 aGIGlyPhePheIleLeuLeuPheLeuLeuThrLeuLeuThrIleuPheThr 203
    |||:::|||||

```

seq_name: sp_organelle:021173

seq_documentation_block:

PRELIMINARY; PRT; 378 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE CYTOCHROME-B.
OS *Heliohobius argenteocinereus* (Silvery mole-rat).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriocognathi; Bathyeridae; *Heliohobius*.

RN [1]

RX MEDLINE; 97370639.

RA FAULKES C.G., ABBOTT D.H., O'BRIEN H.P., LAU L., ROY M.R., WAYNE R.K.,

RT "Micro- and macrogeographical genetic structure of colonies of naked

RL MO1. ECOL. 6:615-628(1997).

$$-i - \text{CATALYTIC ACTIVITY: } QH(2) + 2 \text{ FERRICYTOCHROME C} = Q + 2$$

CC FERROCYTOCHROME C.

CC -!- COFACTOR: TWO HEME GROUPS

CC (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN

CC
CC
(BY SIMILARITY).

DR EMBL; U87527; AAC53270.1; -.

DR PFAM; PF00032; cytochrome_b_C; 1.

DR PFAM; PF00033; cytochrome_b_N; 1.

KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;

| | | | | | | | |
|----|----------|-----|----|-------|----|----------|--------|
| SO | SECURITY | 370 | AA | 42000 | WT | 03003000 | 000000 |
| KW | Heme. | | | | | | |

SEQUENCE 378 AA; 42889 MW; C32032CF CRC32;

alignment_scores:

| | | |
|---------------------|--------|--------------------------|
| Argument Scores: | | |
| Quality: | 69.50 | Length: 102 |
| Ratio: | 1.264 | Gaps: 6 |
| Percent Similarity: | 53.922 | Percent Identity: 29.412 |

DR EMBL: U35170; AAC52540.1; .
DR PFAM; PF00032; cytochrome_b_C; 1.
DR PFAM; PF00033; cytochrome_b_N; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
SQ SEQUENCE 379 AA; 42948 MW; 4FCA3707 CRC32;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 66.50 | Length: | 83 |
| Ratio: | 1.663 | Gaps: | 4 |
| Percent Similarity: | 48.193 | Percent Identity: | 27.711 |

alignment_block:

US-09-030-606-224/rev x Q36033 ..

Align seg 1/1 to: Q36033 from: 1 to: 379

```
217 TTTAGGGCTCCATTTYGAGAAAGTAGATTTCTCTCTCTCTCTTT 168
    ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 PhePheAlaPheHis.....PheIleLeuProPheIleI 189
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
167 AGACATATGATAAATATGCCCCCTTATCTTAAACCATGATCCAT 118
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
189 eThAlaMetVal.....AlaIleHisL 197
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
117 TTTTGTGATGGGAGTAGGGGTACACACT..... 86
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
197 eulePheProHisGluThrGlySerAsnAsnProSerGlyLeuAsnSer 213
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
85 .....CCCTACTACAATGTCACAACTTTT 60
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
214 AspSerAspLysIleProPheHisProTyrTyrThrIleLysAspPheLe 230
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
59 CTTTGTGTTCTTATTTGTTGTAACGTGA.....CTATTCTTCT 20
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
230 uGlyLeuLeuPheMetLeuMetAlaLeuLeuThrLeuThrLeuPheSer 246
```

seq_name: sp_organelle:O20615

seq_documentation_block:

ID 020615 PRELIMINARY; PRT; 379 AA.
AC 020615;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE CYTOCHROME B.
OS Cryptomys hottentotus pretoriae.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriocognathi; Batherygidae; Cryptomys.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-585;
RX MEDLINE; 98067811.
RA FAULKES C.G., BENNETT N.C., BRUFORD M.W., O'BRIEN H.P., AGUILAR G.H.,
RA JARVIS J.U.M.;
RT "Ecological constraints drive social evolution in the African
RT mole-rats."
RL Proc. R. Soc. Lond., B, Biol. Sci. 264:1619-1627(1997).
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C - O + 2
CC FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS
CC (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC (BY SIMILARITY).
DR EMBL: AF012236; AAC53480.1; .
DR PFAM; PF00032; cytochrome_b_C; 1.
DR PFAM; PF00033; cytochrome_b_N; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
SQ SEQUENCE 379 AA; 43155 MW; DB827C3C CRC32;

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 66.50 | Length: | 102 |
| Ratio: | 1.231 | Gaps: | 6 |
| Percent Similarity: | 52.941 | Percent Identity: | 27.451 |

alignment_block:

US-09-030-606-224/rev x O20615 ..

Align seg 1/1 to: O20615 from: 1 to: 379

```
230 GTATCAAGCACCCTTAAAGGCTTCATTTYGAGAAAGTAGT..... 189
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
146 IleThrAsnLeuSerAlaIleProTyrIleGlyThrSerLeuValG1 162
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
188 ....ATTTCTCTCTCTCTCT.....TTTA 167
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
162 UTripIleTribGlyPheSerValAspLysAlaThrLeuThrArgPheP 179
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
166 GAACATATGATAAATATGCCCCCTTAT.....CCTTAAACCATGTAT 123
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
179 heAlaPheHisPheIleLeuProPheIleIleThAlaLeuThrMet.Va 195
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
122 CCATTTTGTGTTGATGGGAGTAGGGGTACACACTCC..... 83
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
195 lHisLeuLeuPheLeuHisGluThrGlySerAsnAsnProSerGlyIleA 212
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
82 .....TACTACAATGTCACAAAC 65
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
212 spSerAsnSerAspLysIleProPheHisArgTyrTyrSerIleLysAsp 228
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
64 TTTTCTTTTGTGTTCTTATGTTGTAAC.....GTACTATT 24
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
229 PheMetGlyPheLeuPheMetPheIleIleLeuAlaPheValLeuPh 245
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
23 TTCT 20
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
245 eSer 246
```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:54 : Search time 1809.22 Seconds
(without alignments)
348.886 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320

Sequence: 1 CCCCTGAGGCTTCTGTGA.....TTTARACTCMGCATGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 2546578 segs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|--------------------|
| 1 | 202.2 | 63.2 | 512 | 21 | R20590 | R20590 y949d06.r1 |
| 2 | 201.2 | 62.9 | 285 | 40 | Z45397 | Z45397 HSC20D041 n |
| 3 | 165.4 | 51.7 | 342 | 32 | AA334754 | AA334754 EST39036 |
| 4 | 152 | 47.5 | 177 | 25 | N56292 | N56292 J55699F Hum |
| 5 | 136.8 | 42.8 | 315 | 32 | AA373221 | AA373221 EST85227 |
| 6 | 40.4 | 12.6 | 214 | 43 | AU033864 | AU033864 AU033864 |
| 7 | 38.4 | 12.0 | 700 | 48 | AI557583 | AI557583 pt2.1-1.D |
| 8 | 37.2 | 11.6 | 310 | 40 | C93124 | C93124 C93124 Dict |
| 9 | 36.8 | 11.5 | 249 | 41 | AI054289 | AI054289 q176c04.x |
| 10 | 36.8 | 11.5 | 269 | 42 | AI144030 | AI144030 q163a08.x |
| 11 | 36.6 | 11.4 | 496 | 23 | HI7024 | HI7024 ym39d10.s1 |
| 12 | 36.6 | 11.4 | 434 | 23 | L47084 | L47084 SCMRAP171 C |
| 13 | 36.6 | 11.4 | 300 | 35 | C57199 | C57199 C57199 Yuj1 |
| 14 | 36.6 | 11.4 | 300 | 36 | C57822 | C57822 C57822 Yuj1 |
| 15 | 36 | 11.2 | 300 | 36 | C59863 | C59863 C59863 Yuj1 |
| 16 | 35.8 | 11.2 | 354 | 44 | AU038221 | AU038221 AU038221 |
| 17 | 35.2 | 11.0 | 504 | 35 | AA550294 | AA550294 1423m3 gm |
| 18 | 35 | 10.9 | 587 | 50 | AI667004 | AI667004 fc24d11.y |
| 19 | 34.8 | 10.9 | 471 | 46 | AI404883 | AI404883 GH24733.5 |
| 20 | 34.6 | 10.8 | 432 | 36 | AA604491 | AA604491 no73a09.s |
| 21 | 34.4 | 10.8 | 305 | 37 | AA721414 | AA721414 nz74b11.s |
| 22 | 34.4 | 10.8 | 520 | 39 | AA825604 | AA825604 od57h06.s |
| 23 | 34.4 | 10.8 | 312 | 43 | AI239707 | AI239707 qh43b02.x |
| 24 | 34.4 | 10.8 | 568 | 44 | AU039059 | AU039059 AU039059 |
| 25 | 34.4 | 10.8 | 384 | 49 | AI630952 | AI630952 ty99g12.x |
| 26 | 34.2 | 10.7 | 442 | 26 | W47600 | W47600 zc35g07.r1 |
| 27 | 34.2 | 10.7 | 323 | 34 | C24351 | C24351 C24351 Dict |
| 28 | 34.2 | 10.7 | 171 | 40 | C90553 | C90553 C90553 Dict |
| 29 | 34.2 | 10.7 | 412 | 43 | AI178240 | AI178240 EST221905 |
| 30 | 34.2 | 10.7 | 406 | 51 | AI705892 | AI705892 UI-R-AC1- |
| 31 | 34 | 10.6 | 543 | 34 | C24372 | C24372 C24372 Dict |
| 32 | 34 | 10.6 | 699 | 45 | AU001859 | AU001859 AU001859 |
| 33 | 34 | 10.6 | 612 | 47 | AI525038 | AI525038 promrna-5 |
| 34 | 34 | 10.6 | 373 | 48 | AI582163 | AI582163 tq63b09.x |
| 35 | 34 | 10.6 | 477 | 53 | HSM000181 | AI035888 Homo sapi |
| 36 | 33.8 | 10.6 | 464 | 20 | Z32876 | Z32876 HHEA41N Atr |
| 37 | 33.8 | 10.6 | 466 | 22 | R22059 | R22059 yh25g02.s1 |
| 38 | 33.8 | 10.6 | 583 | 25 | N50519 | N50519 yy89b04.s1 |
| 39 | 33.8 | 10.6 | 388 | 25 | N97950 | N97950 1711C3 czap |
| 40 | 33.8 | 10.6 | 383 | 25 | N97969 | N97969 1731C3 czap |
| 41 | 33.8 | 10.6 | 427 | 31 | AA279549 | AA279549 zs86b09.r |
| 42 | 33.8 | 10.6 | 565 | 46 | AI407578 | AI407578 EST235868 |
| 43 | 33.8 | 10.6 | 353 | 47 | AI522468 | AI522468 fb19b08.x |
| 44 | 33.8 | 10.6 | 312 | 49 | AV020572 | AV020572 AV020572 |
| 45 | 33.6 | 10.5 | 602 | 34 | AA532223 | AA532223 CPEST.501 |

ALIGNMENTS

RESULT 1
R20590
LOCUS R20590 512 bp mRNA
DEFINITION y949d06.r1 Soares Infant brain INIB Homo sapiens EST 18-APR-1995
IMAGE:36244 5', mRNA sequence.
ACCESSION R20590
NID 9775371
VERSION R20590.1 GI:775371

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 512)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 321
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RPI
High quality sequence stop: 321.
Location/Qualifiers
1. 512
/organism="Homo sapiens"
/db_xref="GDB:408745"
/db_xref="taxon:9606"
/clone="IMAGE:36244"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGATTGCGCGCCGACGAGATTTTCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 141 a 93 c 111 g 158 t 9 others
ORIGIN

Query Match 63.2%; Score 202.2; DB 21; Length 512;
Best Local Similarity 96.9%; Pred. No. 7.8e-38;
Matches 216; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGGACACAAAAAGA 60
DB 76 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGGACACAAAAAGA 135

QY 61 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 120
DB 136 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 195

QY 121 GGATACATGCTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAA 180
DB 196 GGATACATGCTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAA 255

QY 181 GAGAAATACTACTTCTCRAAA--TGGAAGCCCTTAAAGTG 221
DB 256 GAGAAATACTACTTCTCRAAAATGGAAGCCCTTAAAGGG 298

RESULT 2
245397
LOCUS 545397 285 bp mRNA EST 01-MAY-1998

DEFINITION HSC2OD041 normalized infant brain cDNA Homo sapiens cDNA clone
c-2od04, mRNA sequence.
ACCESSION 245397
NID 9574610
VERSION 245397.1 GI:574610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 285)
Aufiray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Marlage-Samson, R., Pletu, G., Pouliot, Y.,
Sebastiani-Kabakchis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
On Sep 12, 1996 this sequence version replaced gi:1407250.

TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read;
Genexpress_library_idt: C;
Genexpress_sequence_idt: Y1c-2od04;.

FEATURES
source
1. 285
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-2od04"
/clone_lib="normalized infant brain cDNA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type-total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaes, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 101 a 44 c 52 g 86 t 2 others
ORIGIN

Query Match 62.9%; Score 201.2; DB 40; Length 285;
Best Local Similarity 98.0%; Pred. No. 1.4e-37;
Matches 200; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACAACCAATAGGACACAAAAAGA 60
DB 82 CCCCTGAAGGCTTCTGTAGAAATAGTACAGTTACAACCAATAGGACACAAAAAGA 141

QY 61 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 120
DB 142 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 201

QY 121 GGATACATGCTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAA 180
DB 202 GGATACATGCTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAA 261

QY 181 GAGAAATACTACTTCTCRAAAT 204
DB 262 GAGAAATACTACTTCTCRAAAT 285


```
RESULT 3
AA334754 342 bp mRNA EST 21-APR-1997
LOCUS EST39036 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA334754
ACCESSION g1987018
VERSION AA334754.1 GI:1987018
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 342)
Adams,M.D., Kerlavage,A.R., Fieischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.R.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280
JOURNAL
MEDLINE
COMMENT On May 9, 1995 this sequence version replaced gi:802598.
Other_ESTs: THC168716
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/cdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
1..342
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):136461"
/db_xref="taxon:9606"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 93 a 66 c 71 g 109 t 3 others
ORIGIN
Query Match 51.7%; Score 165.4; DB 32; Length 342;
Best Local Similarity 96.4%; Pred. No. 2.8e-29;
Matches 163; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 152 TTTATCATATGTTCTAAAGAGAGAGAGAAATACTACTTCTCRAAATGGAAGCC 211
|||||
Db 1 TTTATCATATGTTCTAAAGAGAGAGAGAAATACTACTTCTCRAAATGGAAGCC 60
212 CTTAAGGTGCTTGATCTGAAGGACACAAATGTGGCCGTCCATCCTTTARAGTTG 271
|||||
Db 61 CTTAAGGTGCTTGATCTGAAGGACACAAATGTGGCCGTCCATCCTTTAGAGTTG 120
272 CATGACTTGACACGGTAAGTGTGAGTTTARACTCMGCAATGTGAC 320
```

```
Db 121 CATGACTTGACACGGTAAGTGTGAGTTTARACTCMGCAATGTGAC 169
|||||
RESULT 4
N56292 177 bp mRNA EST 20-FEB-1996
LOCUS J55699F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
DEFINITION clone J55699 5', mRNA sequence.
ACCESSION N56292
VERSION g1199140
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 177)
Llew,C.C.
CDNAs from fetal heart
unpublished (1995)
JOURNAL
MEDLINE
COMMENT On May 18, 1995 this sequence version replaced gi:811148.
Contact: Llew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: llew@cc.utoronto.ca
Seq primer: TCCAAAGAAATTCGGCAGAG.
FEATURES
Location/Qualifiers
1..177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone_lib="J55699"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
BASE COUNT 69 a 27 c 32 g 49 t
ORIGIN
Query Match 47.5%; Score 152; DB 25; Length 177;
Best Local Similarity 92.9%; Pred. No. 3.9e-26;
Matches 158; Conservative 1; Mismatches 11; Indels 0; Gaps 0;
QY 1 CCCCTGAAGCCTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAGA 60
|||||
Db 8 CCCCTGAAGCCTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAGA 67
QY 61 AAAAGTTGTGACATTTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAT 120
|||||
Db 68 AAAAGTTGTGACATTTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAT 127
QY 121 GGATACATGTTAAAGGATARAAGGCCAATATTTATCATATGTTCTAAA 170
|||||
Db 128 GGATACATGTTAAAGGATARAAGGCCAATATTTATCATATGTTCTAAA 177
RESULT 5
AA373221 315 bp mRNA EST 21-APR-1997
LOCUS EST85227 HSCI172 cells I Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA373221
ACCESSION g2025541
VERSION AA373221.1 GI:2025541
```


| KEYWORDS | EST. |
|--|--|
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| AUTHORS | 1 (bases 1 to 315) |
| ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULDER, R.A., BULT, C.J., LEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WAI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITZMAN, J.L., GEORHAGEN, N.S., GLODEK, A., GNEHM, C.L., HANNA, M.C., HEDBLUM, E., HINKLE, P.S., JR., KELLEY, J.M., KELLEY, J.C., LIU, L.-I., MARMAROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEDNARIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DIMKE, D., FENG, D.-F., FERRIE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCH, C., HUNGTUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASELTINE, W.A., FIELDS, C., FRASER, C.M. and VENTER, J.C. | |
| TITLE | Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence |
| JOURNAL MEDLINE | Nature 377 (6547 Suppl), 3-174 (1995) |
| COMMENT | 96026280 On May 18, 1995 this sequence version replaced g1:811155. Other_ESTs: THC168716 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html) Seq primer: M13 Reverse. |
| FEATURES | Location/Qualifiers 1..315 /organism="Homo sapiens" /db_xref="ATCC (inhost):177582" /db_xref="taxon:9606" /clone_lib="HSC172 cells 1" /cell_type="fibroblast" /cell_line="HSC172 (16PDL)" /dev_stage="fetal" /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI" |
| BASE COUNT | 86 a 65 c 64 g 99 t 1 others |
| ORIGIN | |
| Query Match | 42.8%; Score 136.8; DB 32; Length 315; |
| Best Local Similarity | 96.4%; Pred. No. 1.2e-22; |
| Matches 135; Conservative | 4; Mismatches 1; Indels 0; Gaps 0 |
| QY 181 | GAGAAATACTACTTCTCRAAATGGAAGCCCTTAAGGTGCTTGATACGGAAGACAC 240 |
| Db 1 | GAGAAATACTACTTCTCAAAATGGAAGCCCTTAAGGTGCTTGATACGGAAGACAC 60 |
| QY 241 | AAATGTGGCCGTCATCCTCTTARAGTTGCATGACTGGACACGGTAAGTGTTCAGT 300 |
| Db 61 | AAATGTGACCGTCATCCTCTTARAGTTGCATGACTGGACACGGTAAGTGTTCAGT 120 |
| QY 301 | TTTARACTCMGCATTGTGAC 320 |
| Db 121 | TTTGAAGTCAAGCATTGTGAC 140 |

| LOCUS | 214 bp | mrna | EST | 28-APR-1999 |
|-----------------------|---|---|----------------|-------------|
| DEFINITION | AU033864 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium | | | |
| ACCESSION | AU033864 | discoidem cdna clone SLB538, mRNA sequence. | | |
| NID | 93799288 | | | |
| VERSION | AU033864.1 | GI:3799288 | | |
| KEYWORDS | EST. | | | |
| SOURCE | Dictyostelium discoideum. | | | |
| ORGANISM | Dictyostelium discoideum | | | |
| REFERENCE | 1 (bases 1 to 214) | | | |
| AUTHORS | Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mita,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y. | | | |
| TITLE | The Dictyostelium developmental cdna project: generation and analysis of expressed sequence tags from the first-finger stage of development | | | |
| JOURNAL | DNA Res. 5 (6), 335-340 (1998) | | | |
| MEDLINE | 99156227 | | | |
| COMMENT | On May 19, 1998 this sequence version replaced gi:2846208. | | | |
| FEATURES | <p>Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402huesakura.cc.tsukuba.ac.jp PROJECT = 'Dictyostelium discoideum cdna project in Japan' POLYA=No.</p> <p>Location/Qualifiers 1..214 /organism="Dictyostelium discoideum" /strain="AX4" /db_xref="taxon:44689" /clone="SLB538" /clone_lib="Dictyostelium discoideum SL (H.Urushihara)" /dev_stage="slug" 149 a 22 c 12 g 31 t</p> | | | |
| BASE COUNT | 149 a | 22 c | 12 g | 31 t |
| ORIGIN | | | | |
| Query Match | 12.6% | Score 40.4; | DB 43; | Length 214; |
| Best Local Similarity | 51.1% | Pred. No. 2.9; | | |
| Matches | 92; | Conservative 1; | Mismatches 87; | Indels 0; |
| | | | | Gaps 0; |
| QY | 10 GCTTCTGTAGAAATAGTACAGTTACACCAATAGAGAACAAAGAAAGTTTG | 69 | | |
| | 11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 | | | |
| Db | 25 GCCCCTTCTTAATATAAATTGGTCAAAAAAAGAAAAAAGTAAAGTAAAG | 84 | | |
| QY | 70 TGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGGATACATG | 129 | | |
| | 11 | | | |
| Db | 85 AACCAACAATAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAATTT | 144 | | |
| QY | 130 GTTAAAGGATARAAGGCAATATTATCATATGTTCTAAAGAGAGAGAGAAATA | 189 | | |
| | 11 | | | |
| Db | 145 AAATAAAAAGTAAATATTACAAATAAATATCCAAATCCCTAAAAAAAAGAAAAA | 204 | | |
| RESULT | 7 | | | |
| LOCUS | A1557583 | 700 bp | mrna | EST |
| DEFINITION | pt2.1-1.D10.r tumor2 Homo sapiens cdna 3', mRNA sequence. | | | |
| ACCESSION | A1557583 | | | |
| NID | 94489946 | | | |
| VERSION | A1557583.1 | GI:4489946 | | |
| KEYWORDS | EST. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | |
| AUTHORS | Eutheria; Primates; Catarrhini; Homindae; Homo. | | | |
| | 1 (bases 1 to 700) | | | |
| | Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun | | | |
| | Yu,J. and Hood,L. | | | |

| FEATURES | Location/Qualifiers |
|----------|---------------------------------------|
| source | 1. .310 |
| | /organism="Dictyostellium discoideum" |

```

FEATURES
  source
    location/qualifiers
      1. .249
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="8"
        /clone="IMAGE:1862406"
        /clone_1lb="NCI_CGAP_OV26"
        /sex="female"

```



```

/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PAMPI; mRNA made from
papillary serous ovarian carcinoma, cDNA made by oligo-dT
priming. Directionally cloned. Size-selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."
BASE COUNT      80 a      45 c      34 g      90 t
ORIGIN

```

```

Query Match      11.5%; Score 36.8; DB 41; Length 249;
Best Local Similarity 57.4%; Pred. No. 19;
Matches 62; Conservative 2; Mismatches 44; Indels 0; Gaps 0;

```

```

QY  114 AAAAAATGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAGA 173
      ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   70 AAAATATCAAAACTGTTCAGTCATATAAACACAAACATAGTTTATTTAAAAAAA 129
QY  174 GAAGGAAGAGAAAATACTTCTCTCRAAATGGAAGCCCTTAAGGTG 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   130 AAAGAAGAGAAAAATAGTAATTCACAGAAATTCAGTGTTAGACGAG 177

```

```

RESULT 10
A1144030/c 269 bp mRNA EST 29-OCT-1998
LOCUS q163a08.x1 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1861142 3',
DEFINITION mRNA sequence.
ACCESSION A1144030
NID g3665839
VERSION A1144030.1 GI:3665839
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 269)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

```

FEATURES
source
Insert length: 344 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers

```

```

1. 269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1861142"
/clone_lib="NCI_CGAP_Ov26"
/sex="female"
/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PAMPI; mRNA made from
papillary serous ovarian carcinoma, cDNA made by oligo-dT
priming. Directionally cloned. Size-selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."

```

```

BASE COUNT      100 a      36 c      48 g      85 t
ORIGIN

```

```

Query Match      11.5%; Score 36.8; DB 42; Length 269;
Best Local Similarity 57.4%; Pred. No. 19;
Matches 62; Conservative 2; Mismatches 44; Indels 0; Gaps 0;

```

```

QY  114 AAAAAATGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAGA 173
      ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   182 AAAATATCAAAACTGTTCAGTCATATAAACACAAACATAGTTTATTTAAAAAAA 123
QY  174 GAAGGAAGAGAAAATACTTCTCTCRAAATGGAAGCCCTTAAGGTG 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   122 AAAGAAGAGAAAAATAGTAATTCACAGAAATTCAGTGTTAGACGAG 75

```

```

RESULT 11
H17024/c 496 bp mRNA EST 29-JUN-1995
LOCUS ym39d10.s1 Soares infant brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:50782 3', mRNA sequence.
ACCESSION H17024
NID g9883264
VERSION H17024.1 GI:883264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
          Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
          Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
          Trevaskis, E., Waterston, R., Williamson, A., Woldmann, P. and
          Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert size: 1544
High quality sequence stops: 368
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1544 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 368.

```

FEATURES
source

```

```

1. 496
/organism="Homo sapiens"
/db_xref="GDB:423592"
/db_xref="taxon:9606"
/clone="IMAGE:50782"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Iatmid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer (5',
AACTGGAAGAAATTCGGCGCCGACGAATTTTATTTTATTTT 3');
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Iatmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      154 a      95 c      79 g      167 t

```


Wed Sep 29 14:27:50 1999

us-09-030-606-224.rst

Page 9

This Page Blank (uspto)

JOURNAL Science 273 (5278), 1058-1073 (1996)

MEDLINE 96337999

REFERENCE 2 (bases 1 to 10094)

AUTHORS Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghegan,N.S., Weldman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.B., Botodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.

TITLE Direct Submission

JOURNAL Submitted (27-AUG-1996) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

COMMENT On Jan 30, 1998 this sequence version replaced gi:1591965.

FEATURES

source Location/Qualifiers

1. 10094

/organism="Methanococcus jannaschii"

/db_xref="taxon:2190"

95. .364

/gene="MJ1325"

95. .364

/gene="MJ1325"

/note="similar to GB:L10909 SP:P37374 PID:152977 percent identity: 39.74; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="cadmium resistance protein"

/protein_id="AAB99335.1"

/db_xref="PID:g1592329"

/db_xref="GI:1592329"

/translation="MEKYEKAEIEFKAFGDPTRLMLKLLAENGSMCVCKIIDELKKP QPTSHLNLKKGIVKARKEGTWNFYIVDDRVEIILVDEL"

564. .1739

/gene="MJ1326"

564. .1739

/gene="MJ1326"

/note="similar to SP:P17103 PID:43450 percent identity: 50.68; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="GTP-binding protein, member of GTP1/OBG-family"

/protein_id="AAB99336.1"

/db_xref="PID:g1591967"

/db_xref="GI:1591967"

/translation="MLFLSLKRIFFRNHKNFIVGDMGIEEIRRIEELKKTPTYN KATQKHIGRLAKLAKLREQASRGGGGKGIYAKSGDATAFVGPSTYKSTLLN KLTNAKSEGVAYFTLTITVIGILEYKAKIQLDAPGIIIVGASSGKRGTEVLSAVR SADLLITVDIYITLDHLVLEKELYNVGIQLDTPDPVKIKVKEGGINVSSTPLTH IDEDTIEAILNEYRIHNAVYIREDTLEQFIDVAGNRYIIPSLVANKIDLADDEY LKTIKQLEEFKDYILVSGNKGINDLLKEKIYKIGFIKILKPGKKPDEDEPLI MRGATVKDVCCEKLHKDFVNRFRYAQVWGKSAKHPGQRVGLDHLKLEDGILITVIKR"

1748. .2485

/gene="MJ1327"

1748. .2485

/gene="MJ1327"

/note="invalid gene; identified by Genemark; putative"

/codon_start=1

/transl_table=11

/product="M. jannaschii predicted coding region MJ1327"

/protein_id="AAB99343.1"

/db_xref="PID:g1591968"

/db_xref="GI:1591968"

/translation="MVVKLVNNEKLKLSGLRDIREFRIYKIKNDENINDRNLDDKVL SLKDYTISEENLKIKFSEPKDEIYSYEGRTPYDLDCYLINGKNFLIFINKEGDL KSNTRNDVTTYNNLRLYLGIKQRLTSEITINGELVYNNRISGNEIYSGLFVVDKYR RGYKFFLEEIKDDFYVNPNNMFQIRYSPNLGDPIDYFAFVKLLIDALESLSEKSLN SIKTEILVLSIKIQLINIKEGKHGD"

3148. .4044

/gene="MJ1328"

3148. .4044

CDS,

/gene="MJ1328"

/note="similar to GB:L42023 SP:P44414 PID:1003910 PID:1222446 PID:1204764 percent identity: 30.63; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="modification methylase, type II R/M system"

/protein_id="AAB99337.1"

/db_xref="PID:g1591969"

/db_xref="GI:1591969"

/translation="WQKNQTIFLNTWKFLITLHOKPWSFTFTKIPDPHYLLKDIKAV GVLVSGFDEAFLLNEDDISKLNEDERQILKNFVKAKNCKRFYEGFVOYILIEDNLK DEEIFTKYPNYIKLLKLFKDRMENRYLPKNKKFVNWQALRNKFLIKNLKKRIYV TLDKRPYNRFSGLDELPSGDVIFLOPYNDDIYFLGLYLNSSFRRNYLANGARG GRVAFQKLENAKIPFESDEVKEKIKNIYKDIYLNKNGDIENLERQIDYIISAI NNNQFKGYQTTLKNNLKPCLKG"

4049. .4933

/gene="MJ1329"

4049. .4933

/gene="MJ1329"

/note="similar to GB:M26978 SP:P22624 PID:551896 percent identity: 50.90; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="methionyl aminopeptidase (map)"

/protein_id="AAB99338.1"

/db_xref="PID:g1591970"

/db_xref="GI:1591970"

/translation="MEIEGYEKIIEAGKIASKVRREAVKLIXPGVKLLEVAEFVENRI RELGEPAPFCNISINEIAAHYTPKLDNLEFKDDVVKLDGAHVDGYIADTAITVD LNSYKDLVKASEDALYTVIKEINPNMIGEMSKIIQEVIESGYKPISNLSGHVHR YELATGISIPNVERTNOYIDVGDVLAIEPFATDGFQVNGDNLGNIYKFLAKRPITL PQARKLDVISKNPYLPFAERWVLNKNSERLALNSLRASCIIYGYPIIKERENGIVG QAEHTILITENGVEITTK"

5079. .5462

/gene="MJ1330"

5079. .5462

/gene="MJ1330"

/note="similar to GB:AE00666 percent identity: 54.47; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAB99339.1"

/db_xref="PID:g2826396"

/db_xref="GI:2826396"

/translation="MDAGMDLYIGSNERDRVKEKLEIKQLSNPNVSTLLAIL LDNEGRANLFPNYNEDPNYVYVDEVYGLAIAINEIAGTKAIFNFRFYDAKKPGIIGEL DRKGFMLDPAIAGLLAGCMSKVEE"

complement(5465. .5899)

/gene="MJ1331"

complement(5465. .5899)

/gene="MJ1331"

/note="hypothetical protein; identified by Genemark; putative"

/codon_start=1

/transl_table=11

/product="M. jannaschii predicted coding region MJ1331"

/protein_id="AAB99344.1"

/db_xref="PID:g1591972"

/db_xref="GI:1591972"

/translation="WVASMATYELRIYGNVECAEFIDKVESIGKLLDVNGVYVYKD SVRIANFPNEKRQOLFKEIKLDLEDGGLIKVERIERDLNYYIEFPNGLKISTNE LKEINKLDKTISYLENIEFNALEKQIVSEIRDLKDTFEV"

complement(5908. .7089)

/gene="MJ1332"

complement(5908. .7089)

/gene="MJ1332"

/note="similar to SP:P42942 PID:1165215 PID:1323377 PID:790501 percent identity: 39.69; identified by sequence similarity; putative"

/codon_start=1


```

/transl_table=11
/product="GTP-binding protein, member of GTP1/OBG-family"
/protein_id="AAB9340.1"
/db_xref="PID:g1591973"
/db_xref="GI:1591973"
/translation="MAMIGLVGKPNVGKSTMFNALTKEPAEIGNYPFTIQPNKGIAY
ITSPCKELGVKCNPNRNSKCIDGIRHIVPEVIDVAGLVPGAHEGRGNKFLDLRO
ADAFILVVDASGKTDAEGNPTENDVPEDVKFLINIDMIYISILTKNMDKLARAOQ
EKNIYKALKDQLSGLNIDEDDIKMAIRDMDESPKITEEDLNLAKLRKISKPMIIA
ANKADHPDAEKNIERLKEEKDYIVPTSAEIELAKRAEKAGIIKRENDPEIIDES
RVNEQMRRAFDYIKDFLKYGGTGVOECINKAYFDLDMIVVYPEDEKNSDKQGNV
LPDAFLVKKGSTARDLAYVHTELEGEKFIYAIDAKKKIRVGADYELKNDIIKIVSAA
R"
gene
7182..7436
/gene="MJ1333"
7182..7436
/gene="MJ1333"
/feature="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1333"
/protein_id="AAB9345.1"
/db_xref="PID:g1591974"
/db_xref="GI:1591974"
/translation="MSPKIQEILINELDNLMNRKRYIELVATVEYLNLIEPSKREK
FKELYDAETVEDVYELIKAIKQLGMOGARRYLITLLEGO"
complement(7495..7908)
/gene="MJ1333.1"
complement(7495..7908)
/gene="MJ1333.1"
/feature="Brute Force ORF; identified by Genemark; putative"

Query Match
Best Local Similarity 48.2%; Score 45; DB 2; Length 10094;
Matches 120; Conservative 2; Mismatches 127; Indels 0; Gaps 0;

gene
CDS
5 TGAAGCCTTCTGTGTAAGAAATAGTACGTACCAACCAATAGAACACAAAGAGAAAA
64
Db 3438 TGAAGACAACTTAAGATGAGAAATATTAAACCAATATCCAAACATATATAAAA
3497
QY 65 GTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCAATCAAAAAAATGAT 124
Db 3498 ATTGTGAAGTTAAAGACAGATGAGATACCTTCCAAAAAACAATAATGGTT 3557
QY 125 ACATGCTTAAGCATARAAGGCAATATTTATCATATGTCTTAAGAAGAGAGAGA 184
Db 3558 TAATTGGCAGGCAATGAGAACTATAATTTTAAATAAAAAATCTAAATAAAAAACGAAT 3617
QY 185 AAATACTACTTCTCRAAATGGAAGCCCTTAAGGCTTGTGATCTGAAGACACAAAT 244
Db 3618 ATATGTTCCAACCTTGAACAGAAAACCATATAATAGTTAGTTAGAGAGATGAAT 3677
QY 245 GTGCGCGTC 253
Db 3678 ATTACCTTC 3686

RESULT 2
AC006755 199917 bp DNA HTG 23-FEB-1999
LOCUS Caenorhabditis elegans clone Y40C5, WORKING DRAFT SEQUENCE, 1
DEFINITION
unordered pieces.
AC006755
AC006755
NID 94263183
VERSION AC006755.1 GI:4263183
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditoidea; Rhabditiidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 199917)
AUTHORS Waterston,R.H.
```

```

TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 199917)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 199917: contig of 199917 bp in length.
location/Qualifiers
1. 199917
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y40C5"

BASE COUNT 63887 a 33980 c 34366 g 65836 t 1848 others
ORIGIN

FEATURES
source
1. 199917
location/Qualifiers
1. 199917
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y40C5"

Query Match
Best Local Similarity 13.7%; Score 43.8; DB 34; Length 199917;
Matches 90; Conservative 2; Mismatches 82; Indels 0; Gaps 0;

QY 30 ACAGTTACAACCAATAGAACACAAAGAAAGTTGTGACATGTAGTAGGAGTG 89
Db 127380 AAATTTGAACATATTTTGTATGAAAAATGGCAAACTTGGCAGGTTGGCAAGAGTTG 127439
QY 90 TGTACCCCTTACTCCCATCAAAAAAATGATACATGCTTAAGATARAAGGCA 149
Db 127440 TGTAGCTTTTTCACGTTTAACGAGCTTATATCTGGAATAATGAACATRA 127499
QY 150 TATTTATCATATGTTCTAAAGAGAGAGAGAAATACTTCTCRAA 203
Db 127500 GTTATGCCAANTACTGTAAAAAATNTACTAAANCAATTATCTTGTATAAA 127553

RESULT 3
AE001391/c
LOCUS AE001391 14553 bp DNA INV 06-NOV-1998
DEFINITION Plasmodium falciparum chromosome 2, section 28 of 73 of the
complete sequence.
ACCESSION AE001391 AE001362
NID 93845168
VERSION AE001391.1 GI:3845168
KEYWORDS
SOURCE
ORGANISM
malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 14553)
AUTHORS Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Perlea,M.,
Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum
Science 282 (5391), 1126-1132 (1998)
JOURNAL
MEDLINE 99021743
REMARK Erratum:[published erratum appears in Science 1998 Dec
4;282(5395):1827]
2 (bases 1 to 14553)
REFERENCE Gardner,M.J.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
Medical Center Drive, Rockville, MD 20814, USA
FEATURES
location/Qualifiers
```



```

source
1. .14553
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
complement(202. .6234)
/gene="PFB0380c"
complement(202. .6234)
/gene="PFB0380c"
/note="identified by sequence similarity; putative"
/codon_start-1
/product="phosphatase (acid phosphatase family)"
/protein_id="AAC71865.1"
/db_xref="PID:g3845169"
/db_xref="GI:3845169"
/translation="MLIKQEPREVEKKEEKEKKGAKDKGKDLFSLNKKRREKKESQK
IDRYLLNSCDNSKNSYSCCYLNNECFVKNIISICKKMSYEEFKNTKTYIMRHGART
PKRKINTWPFKEGKGDITFLGFOQSIKVGSEIYLRKTYITFNKLNKTYNKRERGLRIN
KEGYIKRKCDCVKKCKTKLYRKNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GYSYLDISTSFNLENRRKRLHKFPLRDEIYYEYFLKINRSNKHQRKYEIKIKR
RRRNITLKIIWQHLLINKMKIKKNMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
DDNDDNNNNNNNNDDNNNNNNDDNNDDNNDDNNDDNNNNNNDDNNNNNNDDNNNNNN
DETFNNKSFNYADM.LKTYKYYKNILDKKNITYTNKKKELFPMELHMYKKLL
INKMEKNIKKKKKYDKIKLKNKYLCKITNSRCKLTAYGIIIGILGISEIYFF
FFILFRSNYDKTDNNIDITYTRKEKKKCLNKRSCFQWILNRDITSGOYCIDKN
TAPVKNYIIGENLGGNGCGKCGDILRGDILGDLTIGDNNISIPFRSRIEFCOS
KITEDELYIYFNKILKRLQSLDDMYKINHEVKMGNDKVLNYSKKCYDKNDGYSY
PSYKNSNDYKSHYIKKMKANVSQCSNESILKERQENKKKKKKKKKKKMENTFINN
NNLMYNINVEFDLIIENERGNOFFYNNIKKKRQKNEGLEWNVYNIQIYKYLNE
ESKEFKLEFLKNVENIDNFTNSITNINYKYYINMYHRKDCFEKKQJHSEKHMKK
IHLRDKFIEYEKENEIIDNCNNINMDKKKEINNANNMIDNNNEIDMSNFIITYY
YIFLYNMDTYIQFLFYLLKNYIILFSVKAERNSLM.LTKTGNHYIKLRNHI
HNSDYKILNNYKDEIFIVYDITKTWENCMTTDLIDVCKNTKIDLENDIPIL
TNDKEEYHVNNSIISVLKKNHSVYKLLKKNLSILKDLKLCNCFINKNYIHTNY
DKHNKIYQDKIKMTYHPFNHKKNVKIIKKFISADAYIYGVNLNLENPAREKLS
QHPSSIDILIKEYGQNNYIINCEIKKYEEQNNFIKRPNINISGKNSCHNKTSSN
TLOGNDRANILDADEDRILKRNKINONRAKQVONGMTINNSKKRYRNOTEYKKEK
KKKYDQNDQTNQKCAQKNDQTNQKNDQTNQTNQTNQTNQTNQTNQTNQTNQTNQTNQ
YKNIYTCYKLMCKNEYSNKYLSWLSGMSLIDVYINILINVLRYEKYNEKTKTCFI
PRIILYTHQSSISLFSQSCVGIIRKDKMIPPFASFTSLELIHKKKIKMLSNCKDC
SNNEKSYCYSNKYNIMKGEKKKHAASSRVHVNQTDPRDYLSFIYHNNTANIFCCKDD
VWKVRETENEKFEKCRKNKKKMEENENYIKDDEKNIYNILKRNINENIDKKSINI
NTCIYNDIPTNVNNKKYESYLPRCLNKIHDFKNLFYLCYKNNNIQDLIQLDCLNN
NYTHIKNMQLKEGKHGKRGFYGVKFTFNNSVPLKLNKLIKKNYNGNKKDKEE
DNNYHNDKNYSNFIYDNRHTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NANKRESLKKKKKKKKKNCIOKNNNICERKKSNIHNNSSYIFNTVRFKKDIARIN
TNKKCDENSISCINNMRKRNIFKLNRLINLNFNSNDKMYIYNSTVITYGKNYK
RINKKDVHINNILLHTYKQKKKSTIISSDNNNNNNNAEDISSRKLFEKDIKNT
KQYIYNDHNNINSYDNNINNGLINLNEKKNVLHNECKKKNQIIGYSIKYDKNVSENSC
SDVITSIKDKKIKKRRKKLQKKNYENENIVCLDLISYLLKMLRIYGNPELL"
9488. .10522
/gene="PFB0385w"
join(9488. .9512,10001. .10200,10376. .10522)
/gene="PFB0385w"
/note="identified by sequence similarity; putative"
/codon_start-1
/product="acyl carrier protein"
/protein_id="AAC71866.1"
/db_xref="PID:g3845170"
/db_xref="GI:3845170"
/translation="MFVLSYYVGYSLQILKKKRSNOVFLNRKNDYNLIKKNPSSS
LKSTFDIKRIISKOLSVEDKIQMNSNFTKDLGADSLDELIMALEXFNVTISDQ
DALKINIVQDAIDYIEKNNKQ"
11515. .13011
/gene="PFB0390w"
11515. .13011
/gene="PFB0390w"
/note="identified by sequence similarity; putative"
/codon_start-1
/product="ribosome releasing factor (OO, TP)"
/protein_id="AAC71867.1"
/db_xref="PID:g3845171"
/db_xref="GI:3845171"
/translation="MIFQHLSLHNEKKISMFFYTLILYISHVRIFYDCNLKNEKN

```

| | | | | | |
|-----------------------|--|---|----------------|---------------|--|
| BASE COUNT | 5522 a | 1421 c | 1141 g | 6469 t | YNYVLKNGNNNSPSYIFLNSYRNNNNFLRRKENYQRYLLLYNPKFNDGAKRNSYILY AHSKKNKKNKINHNINITKRETVSRDRDDEDDDDNYDDDDDDNYDDDDDDNYDD DENYDDDDDDNYDDDDDDNYDDDDDDNYDDDDDDNYDDDDDDNYDDDDDDNYDD PYETQSKSRISKDENVAHQNGKNTLSLENYKIKKNANSLETDKKNVKKKQDEMD ETNNKNNKNETNEKDEKEEKDDNDINDNDIMEDYEEDYITEEHIKDLHKICSEK MKNVYELKKESSYRFLNNVNSNDMEFENEKVKINERIYTVKHICHIKKENLFTTPYD PYFVFLYQHEIFKEFDELKFYKDKSLYAIIPISENKNKIKRRIEDSKVTLR TVRKQMDKLEKFKNKIGKDIYFKQKNYIQSIHDQTKNIEKLFADTK" |
| ORIGIN | 5522 a | 1421 c | 1141 g | 6469 t | |
| Query Match | 13.5%; | Score 43.2; | DB 37; | Length 14553; | |
| Best Local Similarity | 57.7%; | Pred. No. 0.91; | | | |
| Matches | 75; | Conservative 1; | Mismatches 54; | Indels 0; | Gaps 0; |
| QY | 38 | AACCAATAGGACACAAAAAGAAAAAGTTTGACATTGTAGTAGGAGCTGTACCCC | 97 | | |
| Db | 6985 | AACAAACAAAAAATAAATAACAAAGTTGAAAGATTATTGGAAGAAAGCTTTAAATT | 6926 | | |
| QY | 98 | TTACTCCCATCAAAAAAATGATACATGGTTAAAGGATARAAGGCAATATTTTAT | 157 | | |
| Db | 6925 | GTAAGAAGGATAAAAAATAATATATATATATATAAAGAAAAAAGCTAAATTTTAC | 6866 | | |
| QY | 158 | CATATGTTCT | 167 | | |
| Db | 6865 | TTTTTTTTTT | 6856 | | |
| RESULT | 4 | | | | |
| AC005140/c | | | | | |
| LOCUS | AC005140 | 194410 bp | DNA | HTG | 01-APR-1999 |
| DEFINITION | | Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 14 | | | |
| ACCESSION | AC005140 | | | | |
| NID | 94558572 | | | | |
| VERSION | AC005140.3 | GI:4558572 | | | |
| KEYWORDS | HTG; HTGS_PHASE1. | | | | |
| SOURCE | malaria parasite P. falciparum. | | | | |
| ORGANISM | Plasmodium falciparum | | | | |
| REFERENCE | 1 (bases 1 to 194410) | | | | |
| AUTHORS | Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W. | | | | |
| TITLE | Plasmodium falciparum 3D7 chromosome 12 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 194410) | | | | |
| AUTHORS | Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W. | | | | |
| TITLE | Direct Submision | | | | |
| JOURNAL | Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology | | | | |
| COMMENT | Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | | | | |
| | On Apr 2, 1999 this sequence version replaced gi:4337162. | | | | |
| | * NOTE: This is a 'working draft' sequence. It currently | | | | |
| | * consists of 14 contigs. The true order of the pieces | | | | |
| | * is not known and their order in this sequence record is | | | | |
| | * arbitrary. Gaps between the contigs are represented as | | | | |
| | * runs of N, but the exact sizes of the gaps are unknown. | | | | |
| | * This record will be updated with the finished sequence | | | | |
| | * as soon as it is available and the accession number will | | | | |
| | * be preserved. | | | | |
| | 1 13954: contig of 13954 bp in length | | | | |
| | * 13955 14154: gap of unknown length | | | | |
| | * 14155 32431: contig of 18277 bp in length | | | | |
| | * 32432 32631: gap of unknown length | | | | |
| | * 32632 105078: contig of 72447 bp in length | | | | |
| | * 105079 105278: gap of unknown length | | | | |
| | * 105279 117575: contig of 12297 bp in length | | | | |
| | * 117576 117775: gap of unknown length | | | | |
| | * 117776 120340: contig of 2565 bp in length | | | | |
| | * 120341 120540: gap of unknown length | | | | |
| | * 120541 132816: contig of 12276 bp in length | | | | |
| | * 132817 133016: gap of unknown length | | | | |
| | * 133017 135389: contig of 2373 bp in length | | | | |

JOURNAL Submitted (12-OCT-1993) J.E. Hodgson, Smithkline Beecham
Pharmaceuticals, Brockham Park, Betchworth, Surrey, RH3 7AJ, UK

FEATURES

source

1. .4013
/organism="Staphylococcus aureus"

/strain="J2870"

/db_xref="taxon:1280"

/clone="pox 301"

<1. .450

/note="ORF B"

/codon_start=1

/transl_table=11

/protein_id="CAA53190.1"

/db_xref="PID:g438227"

/db_xref="GI:438227"

/db_xref="SWISS-PROT:P41369"

/translation="EPEENHVDLILNLTSTIGYKKNDLKNLLRNSYTSKKGILLID
LNPAPFYKNEEOMIFNEKDYEVNQTIKINKIYNSIVSWDYKKNESKDISRELK
FKLYTNDIICSSLYNFTLSNISPYGDMKGNFHEHSPRLIVKITK"

410. .3588

/gene="1les"

410. .415

/gene="1les"

434. .439

/gene="1les"

461. .466

/gene="1les"

/bound_moiety="promoter"

477. .3551

/gene="1les"

/codon_start=1

/transl_table=11

/product="isoleucyl tRNA synthetase"

/protein_id="CAA53189.1"

/db_xref="PID:g581558"

/db_xref="GI:581558"

/db_xref="SWISS-PROT:P41368"

/translation="MTKKYLNTQNEISAFWNTOKIFKKSIDNRKQESFEVYDGPPTA
NGLPHAGHVLGRVTKDLVARLKTGFGFYERKAGWDTHGLPEVELEVEKKIGIKGQDI
EKYGIENFNECKSVFNEKEWRFESKDLGYWVDMDSPTYLENNYIESVMNISTE
HKKGLYKGHKVTPTCTHDOTALSHEVAQGYKNVKDLSAVVKFQLTNSKDTYFLSWT
TTPWTLPANVALAINKDLNYSKIRENEYIILATDLINSITEKEIIDTFSGSMLIN
LKYIIPFESDGLVNAVYVVDGEFVYNSGEGIVHIAPAHGEDDYQLVERDLPLNVI
TREGVYNDREPELVGNRAKNSDIEIKLSKOLLYKKOKEYEHNPICWRCGNPLIYY
AMEGWFIKTTFNEKNEIINNNNNIEFPISHIKEGRMGNEFMNVAVDNIGRNRWGTPLN
VMICDNCHEYAPSSIKDLONNSINKIDEDIELHRPYVDNITLSCPKNGKMSRYEEV
IDVWFDSGSMFPAQHHPEDNQIFNQHPADFIAGVDQTRGWYSLVISTILKKG
SSYKRALSLGHILDSNGKMSKSGVNTINPTELINKYGADSLRWALLSDAPANNKRE
SENIVAQTSKFEIDLNIYKFYNNYKIDHYNPNNETISRNLTLDNALSRNLTLIK
ESNIYVNNYDPTSAARLINEYTNNTSNWYIGDSRGFEWEGISNDKKDAYNTLYELLT
TL SRLVAPFVPEFISEKIHYNLTGKSVHLQDYPOYKESFINQALDEDMHTVIKIVELSR
QARKNADLKIKOPLSKMVIKPNQSLNLFPLNYYSIIKDELNIKIELTDNINDYITY
ELKLNFSYVGPRLGNKTNIQTLIDSLSEYDKSLIESNNFKSLSSDAELTKDPLIK
TLPKDSYQLSESDNCVILLDKNLSPELIREGHARELIRLIQQLRKKKNLPINQRIIDY
IGVTGELLESIKTNKMKFENFVIKNIHLNVIDEYENTIHFNKKEIKISLTY"

3557. .3588

/gene="1les"

3621. .>4013

/note="ORF C"

/codon_start=1

/transl_table=11

/protein_id="CAA53191.1"

/db_xref="PID:g438228"

/db_xref="GI:438228"

/db_xref="SWISS-PROT:P41370"

/translation="MYNFDYSKLPINKIQKIEPIAGGYVNSFSVDASNKKYFLKLOP
NTKSNFEDYELSLKELTDKNIPVQIINKGELDNNSTFLLEFIENGHAYPESTRKLG
KIVANMHKNTNSLNLFGFSHNFNGTIEF"

BASE COUNT

1591 a 506 c 618 g 1298 t

ORIGIN

.Query: Match

12.6%; Score 40.4; DB 1; Length 4013;

Best Local Similarity 52.4%; Pred. No. 4.3;
Matches 86; Conservative 1; Mismatches 77; Indels 0; Gaps 0;

QY 13 TCTTGTAGAAAATAGTACAGTTACACCAATAGGACACAAAGAAAGTTGTGA 72

Db 3671 TATTTTGAATATTTTAAATGGAATATGAAATATCAAAATATACATAGTTATCA 3612

QY 73 CATGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAATGATACATGTT 132

Db 3611 CCTTTGCTTGCGTTTGATTAACCTCATGCGCACCTACTAAATAGAGTGCCATGTTGT 3552

QY 133 AAAGCATARAAGGCAATATTTATCATATGTTCTTAAAGAGAA 176

Db 3551 TTAATATAATAAGGAAATTTTATTCTTTATTATTAAATGAA 3508

RESULT 7

CEP46C3

LOCUS CEP46C3 29777 bp DNA INV 23-NOV-1998

DEFINITION Caenorhabditis elegans cosmid F46C3, complete sequence.

ACCESSION 266563

VERSION 91051319

KEYWORDS 266563.1 GI:1051319

SOURCE HTG; Myosin; Protein kinase.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 29777)

REFERENCE 1 Cottage, A.

AUTHORS Direct Submission

TITLE Submitted (04-NOV-1995) Louis, MO 63110, USA. E-mail:

JOURNAL jes@sanger.ac.uk or rw@nematode.wustl.edu

REFERENCE 2 (bases 1 to 29777)

AUTHORS Wilson, R., Ahnscough, R., Anderson, K., Baynes, C., Berks, M.,

Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,

Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,

Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,

Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,

Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,

O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,

Saunders, D., Showken, R., Smaison, N., Smith, A., Sonhammer, E.,

Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,

Vaughan, K., Waterston, R., Watson, A., Weinstein, L.,

Wilkinson-Sproat, J. and Woldman, P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans

Nature 368 (6466), 32-38 (1994)

94150718

JOURNAL

MEDLINE

COMMENT

For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence &object=F46C3
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F46C3.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F46C3 is at 1 in this sequence. The true
right end of clone F46C3 is at 9379 in
sequence 268107.
The true left end of clone F48C5 is at 29718 in this sequence. The

true right end of clone F08B12 is at 11562 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268104.
The end of this sequence (29718..29777) overlaps with the start of sequence 268107.

FEATURES

source

Location/Qualifiers

1..29777

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

/clone="F46C3"

complement(5728..7331)

gene

/gene="F46C3.2"

CDS

complement(join(5728..5791,5836..6003,6061..6189,6239..6419,6467..6608,7046..7192,7242..7331))

/gene="F46C3.2"

/codon_start=1

/protein_id="CAA91468.1"

/db_xref="PID:e1346869"

/db_xref="GI:3877185"

/db_xref="SPTREMBL:Q20457"

/translation="MWLTVVGQQLPVKLTLLKGGMLRQQLTIMHVDSDVILDYVTGM DKHQRALLNLIREGVLTDCTINISTHVGPERVOQPMRVPEAMRIIOEGEQLQKE KPMILKIFTMKHYSSASFDNMRADLKVRSLTARYTMPIMWTLFSEFMTLATFASHI WSOVTNYSDLQETLSMTLPITLLVSSLLIGMLLYFNVMVGFLLCCENNSMRLLVTK DLDYSSLYVHFCELNKFEDLRIFYKGLQFILWYTMLLAVYAHEHDLNITYALVCVLT NGFGMLFARMAEFFFRCRRRTDGTSTIAAR"

complement(9727..26596)

gene

/gene="hum-4"

CDS

complement(join(9727..9807,10085..10186,10257..10423,10474..10616,11034..11173,11223..11383,11430..11604,11647..11780,11846..11979,12137..12416,12469..12577,12628..12776,12901..13082,13137..13261,13861..13979,14025..14261,14311..14431,14665..14838,14887..15013,15220..15302,15960..16008,16080..16201,16482..16676,16726..17201,17383..17609,17720..17873,17922..18046,18094..18288,18343..18477,18522..18682,18876..19118,19162..19930,19990..20160,20230..20339,20385..20525,20573..20737,20784..21543,21589..21783,21827..21993,23751..24069,25817..25911,26033..26348,26397..26596))

/note="Similarity to myosin; cDNA EST yk249a4.5 comes from this gene; cDNA EST yk270h6.5 comes from this gene; cDNA EST yk284c3.5 comes from this gene; cDNA EST yk353c4.5 comes from this gene; cDNA EST yk470b4.5 comes from this gene; cDNA EST yk249a4.3 comes from this gene; cDNA EST yk284c3.3 comes from this gene; cDNA EST yk470b4.3 comes from this gene"

/codon_start=1

/protein_id="CAA91469.1"

/db_xref="PID:e1346870"

/db_xref="GI:3877186"

/db_xref="SPTREMBL:Q20456"

/translation="MSTEVIALINPDHHTFDHIEANISRNKNSIACYNNMQTIPPLNFNNSKPAIFKISVKGDLRLPINDDPETATHAPRTNMTAYNPVSTFSRFGCEIA VVNLNCSLKMROHTDQIFENKLVTKLCEVDQRLGSHGNLQSEFPDYLLGIVAN IHFLSEEFYPSSEETLFLSETAEKFKKHVPKRLFECRSSPEIVSLRMCIMDAWVN FGVDMSNFTRLMSDTRWEVETVNOFILRWVRDGRMLDERFELLQSDSTRIGQVSE EMESSTPIHGFEYKEKTSKPSWLASVILKPKQPTKASNVLVFVDFNDKSDQSLW ETSSTGCVNAVAKNALNKIFNNSNAESIVFGESESGSKSYNVEKAFKYLTSQPSKY STKHSSSIEFVFKSGCAKTLKDEATRFGCISIDLKRNVLTLGNLKYTYPLEVPRV ISOKPGERNFENIEFYHGLSDEKAKFGIKGLQKFYINQGNSENIOHDVNRFKHL ESALHYLGFSDDHCMISYKITSITLIHIGNIYFRTKRNPVEQDVEIGNLAEIKLIAF LLEVDFDLVFLPTSEDSQTEJLNALDNDSFAMAYEELFKWNLNRIGLOLKS LHTGVISILDHYFEKYNNGVVEEFLINSYNERIENTLVKHCFFHDLIDYADGISVD YKVPNSIENGKTVELLFKKPYGLLSLTDECKEPKGTHEYLYLHCNLTNHTDSAYGA RNKERLEFGVNRHCIGTWNVNDVEFARNKRIISLSAVOLMNSKNPIGLLESYSGN TSDIIVSAQFVLGAQDIADKINSHVHFVRCIKSNNEOSKFDIPLVNRQIKNLL LAELLSEFVKGYPVKISKITFARYRCLLPGDIAOCONEKEIIODILQGVYKEDDF KIGTEYVFLRERLAERYDGLQNTICGDAIILIQNMKSFYAQVYKRMRAITIKLOSGLRGWKARDYIIRKREMFKAIGRTMKRNKRLDAYHQALGTENSGLOQTIVGYIDINE

CDS

gene

DAKFLERPTSDSDATETLTHYLTVPKKNLSNMNSITLQFAEENFKHLLPRREP IMPELHRESDYDFRSLVEIFKLLIKYMNIDIKLTKKOREDIGRAYIOQISNPCORDE ILVQITINOINKNODKTASDNGKLVHMAISVPEPTENIIPMLIGFNKESVPMKEOLF ATLQRLKIYDSEIARELPNSNLELIATPNTPTVAEINCICYDGLCOVHLSPWSTSE IAEIRLKQGVGIALGWTVEVEPTNRVFAPGTGNHIDVFSQLGANNDSNQNIFF NPPPEKLVKPPVVPKREVYQOEIVSQNTNGTIPKARTYPTPEENTNQPORTESRN EHGKDYDQFLDAVFQYLSKDEQRAAHFNSNQLANTIKGKLARPEGYEPPVOTYSVP PPRPYTLRRVDDSPILRSRAKSLPRIISPRHEHFVRPHSRNSYSNESTSSDDQMYRT RSRERSLPRFHSNNGYNDPSQPYVMMPVQMGHGMILLSPVASEHQAOSRHRTDR SHRHVPSGVERRYMAKRSVDILKPPMSRRTPDAMVEQTYVRPHLAQSPVGRQTSRF EERSALPRGDSRAREMENPDLARQIPRAYPYONGNVPPPPSSYRAPSPAPTSGDR GLNRLPOESYTEPAVKNTKNNSEYLSPHRFENDEOKAVIRHEKETAKNAIMLSDLRL SPVPQPTPPPPPPVREWEVIRDSVERDPVTOGIRGSEFRKEPLVPQPPRPVVAEKP AVKRYKAPWKLTIRKMEFYRGEVLNDIQIIDQYEAQIVEDCKRSYPRLREODRKOVE TVLRQYEIPPSDLNNQSNIHDPDKVAVIETLARLMPLYFNQYVEVEKERDESVSTIFA ISEHGIRLIVHTPHDLENLKIQDFPEFTIADVLSLEANDILSVHVRHDEENAYSAY RIKTNQAPQIKKTLTDLRCLSGGVPRKRFVTLARTGRETPPVGNMWLYGKIENRFGFLAQ YVDSITDGDNVPPPIRHETSEDRDERVRFDFDEVFPSSERYTMIDFATKYFRKPKDKKQ EFWAMEIDISOIVRFSEKPSISQSLADIAGNEESKAVETEHAIMKFMGDEPLKSESMT DVVFKVLLICHRQPDLEDEYCOLIKOTTSNISOKPNKSLRAWRLTLITAYPPSSLT LKPYVLQYLGNADPEWRPEHGTARICQTNMIOETKYGGRKVLNLALEQOITVGSVLY TLVCESTDGCQLRRQARYISKDNVSOQLRPTVAEEMIOELCNLNVRSLSHEQOE SLCTYVGKDKHLNCKNDNYLMDIITSESHKRLPQFYLKRTVWVHPLRYDNAAYIDS MFDQYIDDYLRGSLISTNSGLQTLAATTEIIRKLAAYLFLLPDNPKNKNAKTLPOIV PKSVIEPKHRHQEEMVTRISQLKMGFGMRPAKASHFLELLSTWPTFGVLHYRLKS VVENGHQQLPEVILTINKSGIOLQPKSKVEFKERYNQDIVSVESIRKTAVKIVRLVIN TMOGEETLDIKTDEADEISHLIGQYMFVTGAEERGSTEL"

join(complement(3682..3795),complement(3460..3624), complement(2750..3286),complement(2424..2700), complement(2206..2376),complement(1899..2155), complement(1748..1849),complement(1130..1425), complement(709..1079),complement(385..659), complement(105..304),complement(268104.1:26738..27119), complement(268104.1:26574..26660))

/gene="F46C3.1"

/note="similar to serine/threonine kinase (2 domains); cDNA EST EMBL:D34010 comes from this gene; cDNA EST

EMBL:D36999 comes from this gene; cDNA EST EMBL:D67571 comes from this

gene; cDNA EST yk459d11.3 comes from this gene; cDNA EST yk459d11.5 comes from this gene; cDNA EST yk224a5.5 comes from this gene; cDNA EST yk251c3.5 comes from this gene; cDNA EST yk427c7.5 comes from this gene"

/codon_start=1

/protein_id="CAA91470.1"

/db_xref="PID:e1346871"

/db_xref="GI:3877187"

/db_xref="SPTREMBL:Q19192"

/translation="MSVYIVLAGFLFMALVPENAGQYIDDIIEVSSCQNGYVON AGCDONSQVNIITATLDGVVATLADGETGEMWREDAPLRGTLSSTPIDIGETSL OLMPJLDGRLEFSYTHNMLEPLITTDLSLESTRIGODAVAGKSTYTKGFDLFTG EQRTECSMESCGPGDTEPKNPILIRRTNSTRANDTLGRIEWNLSAETIGVTLG GITSPLTSDVKLLOPPGVIYAVDKNRREEMTNDVGIVSWQVYGNQIGEISIF DPSNIFTTOYEMOREQHNLOQSSLLYMGTSNGEPTIOSP KAKNNL KORANALPEL STMEJLTNPRECTANEERSLAYNVKDETLRLVHNAFRHSQKAIIEKSLSSGAR KLOJIASDTEVSAQRIGTENLSTSVSKSGDGYLVLESEPQVRKFKVNSPITLMQTI ESTIENPTAVVSFLAGLIGTVAVVYINKIAKSPRIEHLSTSTESAETESASHRTYT SFAPTDDEIERFVEEGSDSTPIGAIRHKPLMPRIEKSNIETHTQPOLKPVORLVKTD IDPDEDSFSNDEKRLLRNRTISRSSLEGFTSRFANEFEVKVIGHGGFVVFRAOSI

| | | | | |
|-----------------------|-----------------|----------------|-----------|---------------|
| Query Match | 12.68; | Score 40.4; | DB 36; | Length 29777; |
| Best Local Similarity | 50.0%; | Pred. No. 3.5; | | |
| Matches 95; | Conservative 2; | Mismatches 93; | Indels 0; | Gaps 0; |

| | | | | | | |
|----|------|-------------------|----------------|---------------|----------------|------|
| QY | 43 | ATAGGAACACAAAAAGT | TTGTGACATTTG | AGAGGAGTGTG | TACCCCTACT | 102 |
| | | | | | | |
| Db | 1473 | ACAACAAAAATAGATG | CATACATTTGTTGA | ATTTTAGTTGTTA | TATTAGCGAAAA | 1532 |
| QY | 103 | CCCCATCAAAAAAATG | TACATGTTAAAGAT | ARAAGGCAATAT | TTTATCATAT | 162 |
| | | | | | | |
| Db | 1533 | GTCATTCAAAAATG | CAACGCGTTGAAGA | GAGGGGAATCGT | TTTTGGTTTTTGAA | 1592 |
| QY | 163 | GTTCTAAAGAGAAGA | GAGAAATACTACT | TTCTCRAAATGGA | AGCCCTAAAGTGC | 222 |
| | | | | | | |
| Db | 1593 | CGTATGATCCGAATG | CTTAATAATAAAT | ATTTCAAAATGGA | AATTTCTTAAATTC | 1652 |
| QY | 223 | TTTGATACTG | 232 | | | |
| | | | | | | |
| Db | 1653 | TTGGAAATTG | 1662 | | | |

| | |
|------------|--|
| RESULT | 8 |
| AC007282 | |
| LOCUS | |
| DEFINITION | AC007282 192151 bp DNA Homo sapiens clone NH0484N09, WORKING DRAFT SEQUENCE, 4 unordered pieces. |
| ACCESSION | AC007282 |
| NID | 95001487 |
| VERSION | AC007282.2 GI:5001487 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 192151) |
| AUTHORS | Waterston, R.H. |
| TITLE | The sequence of Homo sapiens clone |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 192151) |
| AUTHORS | Waterston, R.H. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (09-APR-1999) Genome Sequencing Center, Washington |

JOURNAL Submitted (09-APR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Jun 5, 1999 this sequence version replaced gi:4580496.

On Jun 5, 1999 this sequence version replaced gi:4580496.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

| FEATURES | source |
|----------|---------------------|
| * | 1 |
| * | 5125 |
| * | 5142 |
| * | 32145 |
| * | 32162 |
| * | 105761 |
| * | 105778 |
| | location/Qualifiers |
| | 1..192151 |

| | | |
|------------|---------------------------------|-----------|
| | /organism="Homo sapiens" | |
| | /db_xref="taxon:9606" | |
| | /clone="NH0484N09" | |
| BASE COUNT | 62623 a 38154 c 35320 g 56003 t | 51 others |
| ORIGIN | | |

| | | | | |
|--------------------------|--------|----------------|-----------|----------------|
| Query Match | 12.5%; | Score 40; | DB 35; | Length 192151; |
| Best Local Similarity | 50.0%; | Pred. No. 3.6; | | |
| Matches 94; Conservative | 2; | Mismatches 92; | Indels 0; | Gaps 0; |

| | | | |
|----|-------|--|-------|
| QY | 29 | TACAGTTACAACCAATAGGAACACACAAAAAGTTTGTCACATTTGTAGTAGGAGT | 88 |
| | | | |
| Db | 75422 | TAGAGTAAAGACTGTCTCTCAAAAAAAAAAAAAAAAAAGACTTATGAGAGCAGTTAATATA | 75481 |
| QY | 89 | GTGTACCCCTTACTCCCATCAAAAAAAAAAATGGATACATGGTTAAAGATTARAAGGCA | 148 |
| | | | |
| Db | 75482 | TTTAAGTCGACAGATTAAAGTTAAAGTAAGATTCAATGTACTACTACAGAAACCACCA | 75541 |
| QY | 149 | ATATTTCATCATATGTTCTAAAGAGAGAGAAAGAAAATACTCTTCRAATGGAA | 208 |
| | | | |
| Db | 75542 | AACAACATATAATATAATAGTTAAGGAGAGAAAGGAACACAAAGTACACACAAAAACAGAAA | 75601 |
| QY | 209 | GGCCCTTAA | 216 |
| | | | |
| Db | 75602 | TTAATTAA | 75609 |

| | |
|------------|--|
| RESULT | 9 |
| AF129078/c | |
| LOCUS | |
| DEFINITION | AF129078 Homo sapiens chromosome 21q22.1 PAC B22238, complete sequence |
| ACCESSION | AF129078 |
| NID | g4337089 |
| VERSION | AF129078.1 |
| KEYWORDS | HTG. |
| SOURCE | human. |

| ORGANISM | REFERENCE |
|---|---|
| Homo sapiens | 1 (bases 1 to 78338) |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | Relchwald, K., Zimmermann, W., Delabar, J., Dagand, E., Nordsiek, G., |

TITLE Direct Submission
JOURNAL Submitted (17-FEB-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES Location/Qualifiers

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="PAC B22238"
complement(1..153)
/rpt_family="MLT1A1"
/evidence=not_experimental
complement(332..407)
/note="MZEI, score = 83%"
/evidence=not_experimental
1315..1687
/rpt_family="THE1C"
/evidence=not_experimental
1817..1932
/note="GRAIL, score = 59.000%, comment = good"
/evidence=not_experimental
1850..1928
/rpt_family="MLT1J"
/evidence=not_experimental
1860..1914
/note="Xpound exon prediction, score = 65% (0%)"
/evidence=not_experimental
complement(1933..2079)
/rpt_family="MLT1A1"
/evidence=not_experimental
complement(2080..2341)
/rpt_family="AluUc"
/evidence=not_experimental
complement(2342..2549)
/rpt_family="MLT1A1"
/evidence=not_experimental
2721..3167
/rpt_family="LTR1"
/evidence=not_experimental

```


exon complement(2762. .2844)
/note="MZEf, score = 60.4%"
/evidence=not_experimental
repeat_region 3260. .3470
/rpt_family="LTR28"
/evidence=not_experimental
complement(3471. .3538)
/rpt_family="MLT1A2"
/evidence=not_experimental
repeat_region 4631. .4700
/note="homology = 70.00%, score = 21, counts = 10"
/rpt_type=tandem
/rpt_unit=aaaagaa
/evidence=not_experimental
complement(4825. .4872)
/note="GRAIL, score = 84.000%, comment = excellent"
/evidence=not_experimental
exon 5249. .5414
/note="Genscan, score = 4.31%, comment = Terminal_exon 166
bp frame: 2 phase: 1"
/evidence=not_experimental
repeat_region 6153. .6187
/rpt_family="L2"
/evidence=not_experimental
repeat_region 6600. .6701
/rpt_family="L2"
/evidence=not_experimental
repeat_region 6723. .7190
/evidence=not_experimental
/rpt_family="MLT1H"
/evidence=not_experimental
exon 6941. .6992
/note="Xpound exon prediction, score = 66% (0%)"
/evidence=not_experimental
repeat_region 7367. .7665
/rpt_family="L2"
/evidence=not_experimental
repeat_region complement(9762. .9962)
/rpt_family="MER3"
/evidence=not_experimental
repeat_region 9964. .10150
/rpt_family="MIR"
/evidence=not_experimental
repeat_region 10362. .10639
/rpt_family="AluSg"
/evidence=not_experimental
complement(10363. .10619)
/note="MZEf, score = 81.5%"
/evidence=not_experimental
complement(10895. .11136)
/rpt_family="MIR"
/evidence=not_experimental
repeat_region 13416. .13515
/rpt_family="MER5A"
/evidence=not_experimental
repeat_region 14240. .14314
/rpt_family="MER5A"
/evidence=not_experimental
exon 14515. .14671
/note="GRAIL, score = 95.000%, comment = excellent"
/evidence=not_experimental
exon 14515. .14671
/note="MZEf, score = 56.7%"
/evidence=not_experimental
repeat_region 14766. .14791
/note="111 with 96% homology to 112"
/rpt_type=inverted
/evidence=not_experimental
complement(14808. .15551)
/rpt_family="L1PA2"
/evidence=not_experimental
complement(14957. .15105)
/note="GRAIL, score = 63.000%, comment = good shadow"
/evidence=not_experimental

exon complement(15431. .15537)
/note="GRAIL, score = 47.000%, comment = marginal"
/evidence=not_experimental
repeat_region 15552. .15577
/note="112 with 96% homology to 111"
/rpt_type=inverted
/evidence=not_experimental
repeat_region 16981. .17278
/rpt_family="AluSg"
/evidence=not_experimental
complement(17569. .18083)
/rpt_family="MER51B"
/evidence=not_experimental
complement(18364. .18887)
/rpt_family="L1ME2"
/evidence=not_experimental
repeat_region complement(18884. .19040)
/rpt_family="L1ME2"
/evidence=not_experimental
repeat_region complement(19041. .19170)
/rpt_family="FLAM_C"
/evidence=not_experimental
exon 19058. .19149
/note="MZEf, score = 75.9%"
/evidence=not_experimental
complement(19171. .19315)
/rpt_family="L1ME2"
/evidence=not_experimental
repeat_region complement(19314. .19857)
/rpt_family="L1MEC"
/evidence=not_experimental
complement(19837. .19944)
/rpt_family="L1MEC"
/evidence=not_experimental
repeat_region complement(19986. .20285)
/rpt_family="AluSx"
/evidence=not_experimental
exon 20834. .20904
/note="MZEf, score = 52.3%"
/evidence=not_experimental
exon 20834. .20904
/note="GRAIL, score = 70.000%, comment = good shadow"
/evidence=not_experimental
complement(21748. .23646)
/rpt_family="L1PA2"
/evidence=not_experimental
complement(21906. .22046)
/note="GRAIL, score = 69.000%, comment = good shadow"
/evidence=not_experimental
repeat_region 23646. .23832
/rpt_family="L1P"
/evidence=not_experimental
repeat_region 23833. .24258
/rpt_family="LTR7"
/evidence=not_experimental
repeat_region 24267. .24324
/rpt_family="L1P"
/evidence=not_experimental
repeat_region 24325. .24762
/rpt_family="L1P"
/evidence=not_experimental
exon 24341. .24410
/note="MZEf, score = 87.1%"
/evidence=not_experimental
repeat_region 24763. .25071
/rpt_family="AluO"
/evidence=not_experimental
repeat_region 25072. .25617
/rpt_family="L1P"
/evidence=not_experimental
complement(25247. .25096)
/note="GRAIL, score = 71.000%, comment = good"
/evidence=not_experimental

| Query Match | 12.4% | Score 39.8 | DB 42 | Length 78338 |
|-----------------------|--|----------------|---------------|--------------|
| Best Local Similarity | 50.8% | Pred. No. 4.4 | | |
| Matches | 92 | Conservative 1 | Mismatches 88 | Indels 0 |
| | | | | Gaps 0 |
| QY 31 | CAGTTACAACCAATAGGACACAAAAAGTTGTGACATTGTAGTAGGAGTGT | 90 | | |
| Db 36167 | CAGSAAAAAAAAAAAAAAAAAACACAGAGTTGACTATTAAAGCAATCAAGAGGACACCAG | 36108 | | |
| QY 91 | GTACCCCTTACTCCCATCAAAAAAAAAATGATACATGTTAAAGCATARAAGGCAAT | 150 | | |
| Db 36107 | AGAAAGGTGAAGTCCCACTTAAGAGAAAAAAAAAAGATATAGCATATGAATGGAAA | 36048 | | |
| QY 151 | ATTTCATCATATGTTCTAAAGAGAGAAGAGAAATACTACTTCTCRAAATGGAAGC | 210 | | |
| Db 36047 | AATCTCCAGTGAACACAGATAGCATATAAATAAACCAATCACACTTCTCGAAATCAAGGA | 35988 | | |
| QY 211 | C 211 | | | |
| Db 35987 | C 35987 | | | |

| | |
|------------|---|
| RESULT | 10 |
| HS72E17/c | |
| LOCUS | |
| DEFINITION | HS72E17 91029 bp DNA HTG 03-JUN-1999 Homo sapiens chromosome 6 clone 72E17, WORKING DRAFT SEQUENCE, 1n |
| ACCESSION | AL033523 |
| NID | 95002632 |
| VERSION | AL033523.5 GI:5002632 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | human. |

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 91029)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 91029)
Milne, S.
Direct Submission
Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humnyd@anger.ac.uk Clone request: clonerequest@anger.ac.uk
On Jun 7, 1999 this sequence version replaced q1:4678503.

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: dj72E17 Contig_ID: 01487 acc=AL033523 Length: 88938 bp Unfinished: dj72E17 Contig_ID: 01515 acc=AL033523 Length: 1291 bp.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

| FEATURES | source | Location/Qualifiers |
|------------|--------|--------------------------------------|
| | | 1. .91029 |
| | | /organism="Homo sapiens" |
| | | /db_xref="taxon:9606" |
| | | /chromosome="6" |
| | | /clone="72E17" |
| BASE COUNT | 25175 | a 20307 c 20181 g 24566 t 800 others |
| ORIGIN | | |

| | | | | |
|-----------------------|-----------------|-----------------|-----------|---------------|
| Query Match | 12.4%; | Score 39.6; | DB 34; | Length 91029; |
| Best Local Similarity | 49.0%; | Pred. No. 4.8; | | |
| Matches 99; | Conservative 2; | Mismatches 101; | Indels 0; | Gaps 0; |

QY 7 AAGCTTCTTGTTAGAAATAGTACAGTTACAACCAATAGGACACAAAAAGAAAAGT 66
| | ||| | | | | | | | | | | |
Db 44464 AGGTATTCCATTGAAGATGAGCTCACAGACAAAAATTACGAGGCAGATAAGGAATTCAT 44405

| | | | |
|-----|-------|---|-------|
| OY | 67 | T T G T G A C A T T G T A G T A G G G A G T G T G A C C C C T T A C T C C C C A T C A A A A A A A A A A A T G C A T A C | 126 |
| | | | |
| D b | 44404 | A T C A A A C A C T G G T A C A C A C A A T G T T A C T A A C T G C A C A A T A A A T A T G A G A A A A A T A T A C A A | 44345 |
| OY | 127 | A T G G T T A A A G G A T A R A A G G G C A A T A T T T A T C A T A T G T T C T P A A A G A G A A G A A G A G A A A | 186 |
| | | : | |
| D b | 44344 | A A G A A C C A T G S A T A A C A G A G A C A A T C A G A A A A T A A C T T T T A A C A A A G A A G A A G A A T A G A T | 44285 |
| OY | 187 | A T A C T A C T T T C T C R A A A T G G A A | 208 |
| | | : | |
| D b | 44284 | A T A A T A G A A C A T A A A A A G G A A A | 44263 |

| | |
|------------|--|
| RESULT | 11 |
| PFMAL4P3/c | |
| LOCUS | 312766 bp DNA |
| DEFINITION | HTG |
| | 09-APR-1999 |
| | PFMAL4P3 Plasmodium falciparum chromosome 4 strain 3D7, WORKING DRAFT SEQUENCE, in unordered pieces. |

| | |
|-----------|--------------------------------|
| ACCESSION | AL035476 |
| NID | 94494052 |
| VERSION | AL035476.2 |
| KEYWORDS | HTG; HTGS_PHASE1 |
| SOURCE | malaria parasite P. falciparum |
| ORGANISM | Plasmodium falciparum |

| REFERENCE AUTHORS | TITLE JOURNAL |
|---|--|
| Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 312766) Bowman, S., Churcher, C., Harris, D., Lawson, D., Quail, M. and Barrell, B. | Direct Submission Submitted (09-APR-1999) P.falciparum Genome Sequencing Cons |

TITLE Direct Submission
JOURNAL Submitted (09-APR-1999) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Mar 24, 1999 this sequence version replaced qi:4455722.

On Mar 24, 1999 this sequence version replaced gi:4455722.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

| FEATURES | Location/Qualifiers |
|------------|--|
| source | 1..312766 |
| | /organism="Plasmodium falciparum" |
| | /strain="3D7" |
| | /db_xref="taxon:5833" |
| | /chromosome="4" |
| BASE COUNT | 115941 a 29256 c 29590 g 116336 t 21643 others |
| ORIGIN | |

| | | | | |
|--------------------------|--------|----------------|-----------|----------------|
| Query Match | 12.4%; | Score 39.6; | DB 34; | Length 312766; |
| Best Local Similarity | 55.1%; | Pred. No. 4.2; | | |
| Matches 75; Conservative | 1; | Mismatches 60; | Indels 0; | Gaps 0; |

OY 52 ACAAAAAGAAAACGTTTGTCGACATTGTACTAGGGAGTGTGTACCCTTA CTCCCAATCAA 111
 || | | | | | | | | | | | | | |
Db 94865 ACTAAGAAGAAAGATTGCCCTATAATAATAATATGCTCATACCATATAGTCTTAAATAT 94806

QY 112 AAAAAAAAAATGCATCATGCTTAAAGCATARAAGGCCAATATTTTATCATATGTTCTTAAAA 171
||| ||||| | ||||| | | :||| ||| ||||| | | ||| |||||
Db 94805 AAACAAATATAGAAAATGCATGTAGACGCAAAATCAAAACATGTTATCCTTTTATCGAAAA 947466

```

QY      172 GAGAAGGAGAGAAA 187
          | | | | | | | |
Db 94745 AATATGGAAGAGTCAA 94730

```


RESULT 12
AC007659/c
LOCUS
DEFINITION AC007659 89603 bp DNA PLN 27-MAY-1999
Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence,
complete sequence.
ACCESSION AC007659
NID 94895233
VERSION AC007659.1 GI:4895233
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 89603)
REFERENCE Goodman,H.M., Agayre,F., Belmonte,S., Dolan,M., Kilm,U. and
AUTHORS Morris,J.W.
TITLE Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89603)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT Address all correspondence to:
Howard Goodman
Mass. General Hospital
Boston MA 02114
goodmanemgh.harvard.edu

This BAC was annotated in collaboration with TIGR group.
Contact:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone T14P1 is from Arabidopsis chromosome II and is near the
molecular marker m336.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cds.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cds.dtu.dk/netpgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at.at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.

FEATURES
source
1.89603
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="m336"
/clone="T14P1"
complement(1..539)
/gene="T14P1.1"

misc_feature 1..1868
/note="overlap with BAC clone F4L23
(AC002387:121061..122928)."
complement(1..539)
/gene="T14P1.1"
/note="F4L23.31; identical to GB:AAB82643"
complement(97..501)
/gene="T14P1.1"
/note="unknown protein"
/codon_start=1
/protein_id="AAD32819.1"
/db_xref="PID:g4895234"
/db_xref="GI:4895234"
/translation="MASKALAVTALLITLNLFFFTVSTKCPPTTPKPPKTPKSPK
APAVKPTCPDTLKLGVCA DLGLVNVVGSPPKTPCCTLLGLANLEAAVCLCTALK
ANVLGINLVPIDLTLLNTCGKKVPHGQCS"
733..842
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(791..852)
/note="exon predicted by xgrail, quality marginal"
complement(882..898)
/note="exon predicted by xgrail, quality marginal"
complement(join(1180..1571,1671..1723,2116..2171,
2294..2341,2499..2594))
/gene="T14P1.2"
complement(1180..2594)
/gene="T14P1.2"
/note="F4L23.32; identical to GB:AAB82646. Two ESTs do not
appear to be completely spliced"
complement(join(1352..1571,1671..1723,2116..2171,
2294..2333))
/gene="T14P1.2"
/codon_start=1
/product="putative symbiosis-related protein"
/protein_id="AAD32820.1"
/db_xref="PID:g4895235"
/db_xref="GI:4895235"
/translation="MNKGSIFKMDDEFKRAEAGRIREKYPDRIPVIYEAKSEVP
NIDRKRYLVPSDLTVGFVYIRKRIKLSAEKALFVNDVLPPTGLMSSVYEDKKD
EDGLYITYSSENTFGASSI"
complement(3021..3042)
/rpt_family="AT_r1ch"
complement(3443..3571)
/note="exon predicted by xgrail, quality excellent"
3451..3563
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(4034..4074)
/note="exon predicted by xgrail, quality good"
complement(4109..4140)
/note="exon predicted by xgrail, quality marginal"
4194..4302
/rpt_family="(TAAA)n"
complement(4657..4723)
/note="exon predicted by xgrail, quality good"
complement(5290..5327)
/rpt_family="(TA)n"
complement(5891..5921)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
6344..6406
/note="exon predicted by xgrail, quality marginal"
complement(6361..6412)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
6589..6645
/rpt_family="(GAAA)n"
complement(6722..6814)
/rpt_family="(GAAA)n"
6845..8942
/gene="T14P1.3"
6845..8942
gene


```

CDS
    /gene="T14P1.3"
    6845..8767
    /gene="T14P1.3"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AAD32821.1"
    /db_xref="PID:g4895236"
    /db_xref="GI:4895236"
    /translation="MPLSEFERFQEGVEFGLSSSSFYSDSQKINSNODKTEAKQEDLGY
VVGGLFPEPTSLDALRSPSPPLASSTTTLLSSSHGGGTTVTNTVTAGDDNNNK
CSQMGDLDLDGVLSSASSPGQEOSILRLIMDPSGAFGVDPDFGSGSGSPVAPVSDN
SNLCNFPFQETNPFAELINPNCLEFNPNPLSPPAKRFNSGLHQVPELSDPDG
HDPVROHQFOFPFYHNNQQQFPSSSSSTAVAVMPVPSPGAGDDQSVIIEQLFNAA
ELIGTGNNNGDHTVLAOGILARNHHTLSSNHSKSPFORASHTAEALLSLIHNESS
PLITPENILIRIAARYSFSETSPFLQFVNFTANOSILESCNSESDFRIHIDFDVG
GGOWSSLMOELASGVGRRRRNRASLKLTVFAPPSTVSDEFELFTEENLKTAGEV
KIPFEIELLSVELLNPAYWPLSLRSEKEAIAVNLPVNSVASGILPLILFLKQLSP
NIVPCSDRGCDRNDAPFPNAVHSLQYHTSLLESLDANQDDSSIERNFVQPSIEKL
LMKRHWIERSPWPWRLFTQCGFSPASLSQMAEAQAECLLRNBYRGFHVKEKROSLV
MCWQKRELVTVSAMKC"
repeat_region
    10014..10037
    /rpt_family="AT_rich"
    10070..10118
    /note="exon predicted by xgral, quality good"
    complement(10292..10333)
    /note="exon predicted by xgral, quality
    marginal_shadowexon"
    10433..10467
    /rpt_family="AT_rich"
    10954..11000
    /rpt_family="(CAAAA)n"
    complement(join(<11530..11621,11704..11819,11904..12037
    12119..12243,12366..12556,12650..12939,13131..13475))
    /gene="T14P1.4"
    complement(<11530..123475)
    /gene="T14P1.4"
    complement(join(11530..11621,11704..11819,11904..12037,
    12119..12243,12366..12556,12650..12939,13131..13475))
    /gene="T14P1.4"
    /codon_start=1
    /product="putative phosphatidate cytidyltransferase"
    /protein_id="AAD32822.1"
    /db_xref="PID:g4895237"
    /db_xref="GI:4895237"
    /translation="MTNTNTLFLRHDDSEILIEOENSVEPLFVFAWSGIWRAMATFA
ELVLSRCTPCRSFTRKPLIRPLSLGLRLPGDKPLFRSGLGRISVSRFLTAVAR
AESDQDGDHDSKIDRIHNLQNVEDKOKASOLKRVIFIGIGLPGCVLAGWV
FTVALASSVFIGSREYFELVRSRGIAKGMTPPRYVSRVCSVICALMPILTLFENID
ILVTSAAFVAIAILLVQRGSPRFAQLSSTMFGLEYCGYLPSEFWKRLRCGLAAPALNTG
IGRTWPIILGGQAHWTVGVLATLISFSGVIATDTFAFLGKTFGRPLTISPKTWE
GTVGLVGCIAITILLSKYLSWPOSLEFSVAFFGLNFFGVSFGLDLESIMIKRDAGVKD
SGSLIPHGGLDRVDSYIFTGALASYFKTSLKLYG"
    complement(join(13660..14019,14258..14335,14868..14918,
    14999..15157,15239..15354,15460..15541,15621..15691,
    15800..15877))
    /gene="T14P1.5"
mRNA
Query Match      12.1%; Score 38.8; DB 8; Length 89603;
Best Local Similarity 48.5%; Pred. No. 7.2;
Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0;
QY 33 GTTACAACCAATAGGAACAACAAAAAGTTTGACATTGTAGTAGGAGTGTGT 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84605 GTTGGATCATACACAATGTAGAACAAAAAGAGATTTAATTGAAAAATCAATTGCGG 84546
QY 93 ACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGATARAAGGCAATAT 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84545 AAACCATGTTCCAATTCACAAACAAAAAGAGATATCACCAAGAATTAACCTTAACAG 84486
QY 153 TTATCATATGTTCTAAAGAGAAGAAATACTTCTCTCRAAATGGAAGCCC 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84485 TGTACCTATCTCTATCATCATTAACAAATCCCAAAATGCGTTCCTCAAAAACTCTGATT 84426

```

| QY | 213 | TTAAAGTGTTCATGACTGAAGC | 238 |
|--------------|---|----------------------------|-------------------|
| Db | 84425 | TGCCATATCCAAATTATCTTCATGAC | 84400 |
| RESULT | 13 | | |
| ATAC002388/c | | | |
| LOCUS | ATAC002388 | 85534 bp | DNA |
| DEFINITION | Arabidopsis thaliana chromosome II BAC T13E15 | | 14-DEC-1998 |
| | | | genomic sequence, |
| ACCESSION | AC002388 | | |
| NID | 93420042 | | |
| VERSION | AC002388.1 | GI:3420042 | |
| KEYWORDS | HTG. | | |
| SOURCE | thale cress. | | |
| ORGANISM | Arabidopsis thaliana | | |
| | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. | | |
| REFERENCE | 1 (bases 1 to 85534) | | |
| AUTHORS | Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R. and Venter,J.C. | | |
| TITLE | Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 85534) | | |
| AUTHORS | Rounsley,S.D. and Lin,X. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (25-JUL-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, Email: rounsley@tigr.org | | |
| REFERENCE | 3 (bases 1 to 85534) | | |
| AUTHORS | Rounsley,S.D. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (27-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA | | |
| REFERENCE | 4 (bases 1 to 85534) | | |
| AUTHORS | Rounsley,S.D. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (15-AUG-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA | | |
| COMMENT | On Aug 15, 1998 this sequence version replaced gi:2344885. | | |

e-mail: rounsley@tigr.org
BAC clone T13E15 is from Arabidopsis chromosome II and contains the molecular marker m36.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at/at.html>).

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named as isoforms. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted

| FEATURES | | exons by GRAIL are annotated as misc features. |
|---------------|---|--|
| source | Location/Qualifiers | |
| MRNA | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m336" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLPDVDS SDAGSVNFVFGDCTNPSTVSHEPAIIFS CVDDSGNWGRGMFDALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQONTQA SKPLWAVAVTOSYNSRRKVRSSISIPDLECLAKASFASQKSAJLHMPRIGYQDG SDRSOWTVVERLKRYSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| misc_feature | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGQSRDGEVTPLVGGNGDNVAVAPQVENSIPALNEASYI TQATSYLSLEYGKDTISHPHELLRSTGVGNSPVSVCSISPGERSTSEASTSDA NSPSRESTEILPOATNAIVTSNRLNLNGISMEQGLIERARTVRGSDDIGLQRAPE MPVEDGTDRENFKILEDIGNHGLYFVDTKFKSKMGLACHIAKIHSSSVEKNAREI KEYIEELCWGSNKRVLILGHSGGIDAAALSLYPELKDRAVGLAQSYPGSPIA TDILREGOLGDVNLKRMKIELISKVIGKIDIGLEDLYERRKEFLKNHPLPRELPTV SFRTEASISPAVLSTISHVAHAELPLTNOAKLPVMPPLGAMAACAQLQVRYGEKS DGLVTCDAEVPGSVVVRPKRKLDHAMWVSSLSNEVPLEADAQVCEALLTLVQVEQ ERQOKLATKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grail" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRFSSISPESEAAAARKOPFTPSSSEDKLDTAV LQFQNLKSQKLEAQVECSILEDKLSQIKEROLPYNSLKTIVKSWEKQLTASVESC SVRVSDDSSGAHRFVAKEDGSSPAVKNDFINRLLETGATESSSSNCSNQMEENGVT SSOMTQTLNVLVAATEDLRCIKDELPTVLRNMLGKDLGQALSELSESEIKSFRGDL DVLVFKFSLSRELQSHRDADAKVRVDLKRINGSYLVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| FEATURES | | exons by GRAIL are annotated as misc features. |
|---------------|---|--|
| source | Location/Qualifiers | |
| MRNA | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m336" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLPDVDS SDAGSVNFVFGDCTNPSTVSHEPAIIFS CVDDSGNWGRGMFDALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQONTQA SKPLWAVAVTOSYNSRRKVRSSISIPDLECLAKASFASQKSAJLHMPRIGYQDG SDRSOWTVVERLKRYSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| misc_feature | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGQSRDGEVTPLVGGNGDNVAVAPQVENSIPALNEASYI TQATSYLSLEYGKDTISHPHELLRSTGVGNSPVSVCSISPGERSTSEASTSDA NSPSRESTEILPOATNAIVTSNRLNLNGISMEQGLIERARTVRGSDDIGLQRAPE MPVEDGTDRENFKILEDIGNHGLYFVDTKFKSKMGLACHIAKIHSSSVEKNAREI KEYIEELCWGSNKRVLILGHSGGIDAAALSLYPELKDRAVGLAQSYPGSPIA TDILREGOLGDVNLKRMKIELISKVIGKIDIGLEDLYERRKEFLKNHPLPRELPTV SFRTEASISPAVLSTISHVAHAELPLTNOAKLPVMPPLGAMAACAQLQVRYGEKS DGLVTCDAEVPGSVVVRPKRKLDHAMWVSSLSNEVPLEADAQVCEALLTLVQVEQ ERQOKLATKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grail" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRFSSISPESEAAAARKOPFTPSSSEDKLDTAV LQFQNLKSQKLEAQVECSILEDKLSQIKEROLPYNSLKTIVKSWEKQLTASVESC SVRVSDDSSGAHRFVAKEDGSSPAVKNDFINRLLETGATESSSSNCSNQMEENGVT SSOMTQTLNVLVAATEDLRCIKDELPTVLRNMLGKDLGQALSELSESEIKSFRGDL DVLVFKFSLSRELQSHRDADAKVRVDLKRINGSYLVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| FEATURES | | exons by GRAIL are annotated as misc features. |
|---------------|---|--|
| source | Location/Qualifiers | |
| MRNA | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m336" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLPDVDS SDAGSVNFVFGDCTNPSTVSHEPAIIFS CVDDSGNWGRGMFDALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQONTQA SKPLWAVAVTOSYNSRRKVRSSISIPDLECLAKASFASQKSAJLHMPRIGYQDG SDRSOWTVVERLKRYSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| misc_feature | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGQSRDGEVTPLVGGNGDNVAVAPQVENSIPALNEASYI TQATSYLSLEYGKDTISHPHELLRSTGVGNSPVSVCSISPGERSTSEASTSDA NSPSRESTEILPOATNAIVTSNRLNLNGISMEQGLIERARTVRGSDDIGLQRAPE MPVEDGTDRENFKILEDIGNHGLYFVDTKFKSKMGLACHIAKIHSSSVEKNAREI KEYIEELCWGSNKRVLILGHSGGIDAAALSLYPELKDRAVGLAQSYPGSPIA TDILREGOLGDVNLKRMKIELISKVIGKIDIGLEDLYERRKEFLKNHPLPRELPTV SFRTEASISPAVLSTISHVAHAELPLTNOAKLPVMPPLGAMAACAQLQVRYGEKS DGLVTCDAEVPGSVVVRPKRKLDHAMWVSSLSNEVPLEADAQVCEALLTLVQVEQ ERQOKLATKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grail" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRFSSISPESEAAAARKOPFTPSSSEDKLDTAV LQFQNLKSQKLEAQVECSILEDKLSQIKEROLPYNSLKTIVKSWEKQLTASVESC SVRVSDDSSGAHRFVAKEDGSSPAVKNDFINRLLETGATESSSSNCSNQMEENGVT SSOMTQTLNVLVAATEDLRCIKDELPTVLRNMLGKDLGQALSELSESEIKSFRGDL DVLVFKFSLSRELQSHRDADAKVRVDLKRINGSYLVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| FEATURES | | exons by GRAIL are annotated as misc features. |
|---------------|---|--|
| source | Location/Qualifiers | |
| MRNA | 1..85534 | /organism="Arabidopsis thaliana" |
| | | /cultivar="Columbia" |
| | | /db_xref="taxon:3702" |
| | | /chromosome="11" |
| | | /map="m336" |
| | | /clone="T13E15" |
| | join(<183..313,385..530,611..775,902..>1020) | |
| | /gene="T13E15.1" | |
| | <183..>1020 | |
| | /gene="T13E15.1" | |
| gene | /note="predicted by genefinder" | |
| | join(<183..313,385..530,611..775,902..1020) | |
| | /gene="T13E15.1" | |
| | /note="hypothetical protein, 5' partial" | |
| | /codon_start=1 | |
| | /protein_id="AAC31826.1" | |
| | /db_xref="PID:g2344886" | |
| | /db_xref="GI:2344886" | |
| | /translation="SLVEEPIPLPDVDS SDAGSVNFVFGDCTNPSTVSHEPAIIFS CVDDSGNWGRGMFDALSKLSNTVPTAYHRASEFKDLHLGLHLIKIDNDQONTQA SKPLWAVAVTOSYNSRRKVRSSISIPDLECLAKASFASQKSAJLHMPRIGYQDG SDRSOWTVVERLKRYSIFTVKIFV" | |
| | join(<1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..>4461) | |
| misc_feature | /gene="T13E15.2" | |
| | <1966..>4461 | |
| | /gene="T13E15.2" | |
| | join(1966..2016,2252..2360,2553..2799,2965..3093, 3350..3437,3518..3805,3988..4461) | |
| | /gene="T13E15.2" | |
| | /note="unknown protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31827.1" | |
| | /db_xref="PID:g2344887" | |
| | /db_xref="GI:2344887" | |
| repeat_region | /translation="MGEAGQSRDGEVTPLVGGNGDNVAVAPQVENSIPALNEASYI TQATSYLSLEYGKDTISHPHELLRSTGVGNSPVSVCSISPGERSTSEASTSDA NSPSRESTEILPOATNAIVTSNRLNLNGISMEQGLIERARTVRGSDDIGLQRAPE MPVEDGTDRENFKILEDIGNHGLYFVDTKFKSKMGLACHIAKIHSSSVEKNAREI KEYIEELCWGSNKRVLILGHSGGIDAAALSLYPELKDRAVGLAQSYPGSPIA TDILREGOLGDVNLKRMKIELISKVIGKIDIGLEDLYERRKEFLKNHPLPRELPTV SFRTEASISPAVLSTISHVAHAELPLTNOAKLPVMPPLGAMAACAQLQVRYGEKS DGLVTCDAEVPGSVVVRPKRKLDHAMWVSSLSNEVPLEADAQVCEALLTLVQVEQ ERQOKLATKID" | |
| | complement(4471..4553) | |
| | /rpt_family="(TAAA)n" | |
| | 5103..5185 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| | join(<5213..5329,5510..5671,5752..5833,5953..6180, 6279..>6481) | |
| | /gene="T13E15.3" | |
| | <5213..>6481 | |
| | /gene="T13E15.3" | |
| | /note="predicted by grail" | |
| CDS | join(5213..5329,5510..5671,5752..5833,5953..6180, 6279..6481) | |
| | /gene="T13E15.3" | |
| | /note="hypothetical protein" | |
| | /codon_start=1 | |
| | /protein_id="AAC31828.1" | |
| | /db_xref="PID:g2344888" | |
| | /db_xref="GI:2344888" | |
| | /translation="MASTGEPPDRKRHRFSSISPESEAAAARKOPFTPSSSEDKLDTAV LQFQNLKSQKLEAQVECSILEDKLSQIKEROLPYNSLKTIVKSWEKQLTASVESC SVRVSDDSSGAHRFVAKEDGSSPAVKNDFINRLLETGATESSSSNCSNQMEENGVT SSOMTQTLNVLVAATEDLRCIKDELPTVLRNMLGKDLGQALSELSESEIKSFRGDL DVLVFKFSLSRELQSHRDADAKVRVDLKRINGSYLVSSHSTIYC" | |
| | 6935..6996 | |
| | /note="exon predicted by xgrail, quality marginal" | |
| misc_feature | complement(7199..7378) | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| Query Match | 12.18; | Score 38.8; | DB 8; | Length 85534; |
|-----------------------|-----------------|---|-----------|---------------|
| Best Local Similarity | 48.58; | Pred. No. 7.2; | | |
| Matches 100; | Conservative 2; | Mismatches 104; | Indels 0; | Gaps 0; |
| OY | 33 | GTACAAACCAATAGGAACAACAAAAAGAAAGTTGTGACATTGTAGTAGGAGTGTT | 92 | |
| | | | | |


```

repeat_region 21434. .21497
/rpt_family="(CAAAA)n"
repeat_region 21719. .21743
/rpt_family="(TAAA)n"
repeat_region complement(21747. .21869)
/rpt_family="AT_rich"
repeat_region complement(22138. .22846)
/rpt_family="Lx4"
repeat_region complement(22875. .23251)
/rpt_family="L1_MM"
repeat_region complement(23235. .23520)
/rpt_family="Lx4"
repeat_region complement(23520. .23698)
/rpt_family="L1"
repeat_region complement(23926. .24056)
/rpt_family="Lx3"
repeat_region complement(24353. .24390)
/rpt_family="AT_rich"
repeat_region 24885. .25275
/rpt_family="Lx4"
repeat_region complement(25276. .25307)
/rpt_family="POLY_A"
repeat_region complement(25416. .26656)
/rpt_family="L1_MM"
repeat_region 26657. .28391
/rpt_family="Lx4"
repeat_region complement(28448. .28491)
/rpt_family="AT_rich"
repeat_region 28945. .29032
/rpt_family="AT_rich"
repeat_region complement(29898. .29954)
/rpt_family="(TAAA)n"
repeat_region complement(29957. .30863)
/rpt_family="Lx2B"
repeat_region complement(30905. .32970)
/rpt_family="L1"
repeat_region complement(33018. .33876)
/rpt_family="L1"
repeat_region complement(33895. .33998)
/rpt_family="L1"
repeat_region complement(34617. .34657)
/rpt_family="AT_rich"
repeat_region 35012. .35139
/rpt_family="B1_MM"
repeat_region complement(35253. .36201)
/rpt_family="L1"
repeat_region complement(36211. .37107)
/rpt_family="Lx2B"
repeat_region 37094. .37414
/rpt_family="Lx3"
repeat_region 37417. .37598
/rpt_family="Lx2"
repeat_region complement(37613. .38174)
/rpt_family="L1VL2"
repeat_region complement(38118. .38381)
/rpt_family="L1VL2"
repeat_region 38410. .38945
/rpt_family="Lx7"
repeat_region complement(39080. .39151)
/rpt_family="AT_rich"
repeat_region 39216. .39249
/rpt_family="AT_rich"
repeat_region complement(39401. .39940)
/rpt_family="RMER6A"
repeat_region 41499. .41933
/rpt_family="L1_MM"

```

| | | | | |
|---------------------------|--------|-----------------|-----------|---------------|
| Query Match | 12.18; | Score 38.8; | DB 12; | Length 69037; |
| Best Local Similarity | 48.58; | Pred. No. 7.4; | | |
| Matches 100; Conservative | 2; | Mismatches 104; | Indels 0; | Gaps 0; |

QY 11 CTTCTTGTAGAAATAGTACAGTTACACCAATAGGAACAACAAAAGAAAGTTTGT 70
 ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

| | | | | | | |
|----|-------|-----------------------|-----------------------|-------------------------|----------|-------|
| Db | 65351 | CTTCCTATTAAATPAAACACA | CAATTTTAAACATGAATAG | AAAAAAATCTAGGCA | CATTGTGG | 65410 |
| QY | 71 | GACATTGTAGAGGAGTGT | TACCCCTTACTCCCCATCA | AAAAAAATGATACATGG | | 130 |
| Db | 65411 | TATAGAAAGATCAACAT | AAAAATTCCTAGGGGACTGA | AAAAAATATTAATTTATGCATGT | | 65470 |
| QY | 131 | TTAAAGGATARAAGGCA | TATTTTATCATATGTTCT | TAAGAGAAAGAGAAATAC | | 190 |
| Db | 65471 | ATPAAACATTAGGAGCAG | AAATATTTTACTCATTTTCAT | TCTTAAAGATGACTAATGTCC | | 65530 |
| QY | 191 | TACTTCTCRAAAATGGA | AGCCCTTAA | | | 216 |
| Db | 65531 | TATGCTCTTAAACATPAG | CCCAATTA | | | 65556 |

| | | | | |
|------------|---|---|----------------------|-----------------|
| RESULT | 15 | | | |
| LOCUS | HS800F24/c | | | |
| DEFINITION | HS800F24 | 138888 bp | DNA | PRI 12-MAY-1999 |
| | Human DNA | sequence from clone | 800F24 on chromosome | 1q24, complete |
| | sequence. | | | |
| ACCESSION | Z99758 | | | |
| NID | 94775628 | | | |
| VERSION | Z99758.7 | GI:4775628 | | |
| KEYWORDS | HTG. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; | | | |
| | Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | |
| REFERENCE | 1 (bases 1 to 138888) | | | |
| AUTHORS | Grafham,D. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (12-MAY-1999) | Sanger Centre, Hinxton, Cambridgeshire, | | |

requests: clonerequest@sanger.ac.uk
On May 11, 1999 this sequence version replaced qi:4678515.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em:; EMBL; Sw:; SWISSPROT; Tr:; TREMBL

IMPORTANT: This sequence is not the entire insert of clone 800F24. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone dj1018k9 is at 138789 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> 800F24 is from the library RPC14 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR>: PCYPAC2.

```

FEATURES
  source
    Location/Qualifiers
      1..13888
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="800F24"
        /map="q24"
        /clone_lib="RPC14"
      1..1964
        /note="11 repeat: matches 3306..5335 of consensus"
    repeat_region
    repeat_region
      1967..2267

```


| | | | |
|---------------|---|---------------|---|
| repeat_region | /note="AluSc repeat: matches 1. .300 of consensus" | repeat_region | 27202. .27345 |
| repeat_region | /note="LIP2 repeat: matches 5413. .5484 of consensus" | repeat_region | /note="L2 repeat: matches 2559. .2710 of consensus" |
| repeat_region | /note="LIM1 repeat: matches 799. .1710 of consensus" | repeat_region | 27849. .28052 |
| repeat_region | /note="LIM1 repeat: matches 491. .794 of consensus" | repeat_region | /note="MIR repeat: matches 1. .205 of consensus" |
| repeat_region | /note="MIR repeat: matches 87. .216 of consensus" | repeat_region | 28914. .29215 |
| repeat_region | /note="MIR repeat: matches 49. .146 of consensus" | repeat_region | /note="AluX repeat: matches 1. .310 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .464 of consensus" | repeat_region | 29892. .30582 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP16 repeat: matches 5476. .6155 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 30596. .31979 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP16 repeat: matches 4744. .6155 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 32709. .32764 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 2654. .2709 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 34717. .35145 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1906. .2371 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 35142. .35391 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 2141. .2359 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 35392. .35691 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="AluX repeat: matches 1. .300 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 35692. .37115 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 477. .2141 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 37415. .37828 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="MIR repeat: matches 6. .426 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 38428. .38572 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="MIR repeat: matches 111. .262 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 38626. .39246 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIM4 repeat: matches 4297. .4933 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 39346. .39651 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="Alu repeat: matches 1. .301 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 39699. .40087 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1879. .2298 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 40158. .40665 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIM4 repeat: matches 4796. .5332 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 40666. .40992 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="TIGER1 repeat: matches 2078. .2418 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 41012. .42354 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP2 repeat: matches 4784. .6144 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 43109. .43233 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="MIR repeat: matches 20. .148 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 43333. .43694 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1188. .1584 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 44121. .44349 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="MER8 repeat: matches 8. .239 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 44360. .44465 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 2602. .2712 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 44502. .44695 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 2302. .2491 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 45695. .45961 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="Alu repeat: matches 1. .299 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 46075. .46116 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1. .299 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 46137. .46200 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1. .299 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 46646. .46790 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 1. .299 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 46899. .47049 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="L2 repeat: matches 102. .258 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 47326. .47575 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="MER46A repeat: matches 1. .236 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 47845. .48314 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP repeat: matches 4916. .5395 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 48303. .52827 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP repeat: matches 1236. .5781 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 52826. .57156 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP2 repeat: matches 2. .4320 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 57157. .57376 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="Alu repeat: matches 94. .311 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 57377. .59193 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="LIP2 repeat: matches 4320. .6144 of consensus" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 58567. .58667 |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | /note="Single clone region" |
| repeat_region | /note="LIR16B repeat: matches 1. .309 of consensus" | repeat_region | 59220. .59596 |

Query Match

12.0%; Score 38.4; DB 10; Length 138888;

| | | |
|-----------------------|------------------|----------------|
| Best Local Similarity | 50.0%; | Pred. No. 8.4; |
| Matches | 93; Conservative | 1; Mismatches |
| | | 92; Indels |
| | | 0; Gaps |
| | | 0; |

20 AGAAATAGTACAGTTACAACCAATAGGACACAAAAAGTTGTGACATTGTA 79

Db 125120 AGAAGGAAAAAGAGACAACAAGAAAAACAATAATGATTTTTTCCAAATTCCT 125061

80 GTAGGAGTGTGTACCCCTTACTCCCCATCAAAAAAATGGATACATGGTTAAAGGAT 139

Db 125060 TTAAGGCAATCTTTTCACCTTATACTAGAAATAGTCTGAAATGCAAGTAT 125001

QY 140 ARAAGGCCATATTTTATCATATGTTCTAAAGAGAAGGAGAGAAAACTACTTCTC 199

Db 125000 TTATGAGTTTGGTGTATCAGTATGAATAAGCTTTAATAACAGACAATGTACTTTCTA 124941

QY 200 RAAATG 205

Db 124940 ACATTG 124935

Search completed: September 25, 1999, 23:46:10
Job time: 7913 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:58:34 ; Search time 273 Seconds
(without alignments)
293.265 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320
Sequence: 1 CCCCTGAGGCTTCTGTGTA.....TTTARACTCMGCATTGTGAC 320

Scoring table: IDENTITY_NTUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------|
| 1 | 318 | 99.4 | 320 | 1 | V58694 | Prostate tumour sp |
| 2 | 318 | 99.4 | 320 | 1 | V61269 | CDNA sequence of p |
| 3 | 45 | 14.1 | 110000 | 1 | V21209_01 | Continuation (2 of |
| 4 | 37 | 11.6 | 110000 | 1 | V21209_15 | Continuation (16 o |
| 5 | 36.6 | 11.4 | 19124 | 1 | T72882 | Plasmodium var-7 g |
| 6 | 35.6 | 11.1 | 1215 | 1 | T68041 | H. pylori cytoplas |
| 7 | 35.6 | 11.1 | 717 | 1 | T67711 | H. pylori cytoplas |
| 8 | 35 | 10.9 | 1107 | 1 | V74949 | Staphylococcus aur |
| 9 | 34.6 | 10.8 | 1811 | 1 | V17183 | Insecticidal toxin |
| 10 | 34.2 | 10.7 | 10813 | 1 | V74675 | Staphylococcus aur |
| 11 | 34.2 | 10.7 | 110000 | 1 | X20248_04 | Continuation (5 of |
| 12 | 34.2 | 10.7 | 116277 | 1 | X20249 | Borrelia burgdorfe |
| 13 | 34 | 10.6 | 9842 | 1 | X20259 | Borrelia burgdorfe |
| 14 | 33.8 | 10.6 | 110000 | 1 | X20248_03 | Continuation (4 of |
| 15 | 33.2 | 10.4 | 7295 | 1 | T72883 | Plasmodium var-1 g |
| 16 | 32.8 | 10.2 | 64976 | 1 | V21209_16 | Continuation (17 o |
| 17 | 32.8 | 10.2 | 2120 | 1 | X33812 | Coding sequence fo |
| 18 | 32.6 | 10.2 | 110000 | 1 | X20248_08 | Continuation (9 of |
| 19 | 32.4 | 10.1 | 1406 | 1 | T41619 | Tomato endo-xylogl |
| 20 | 32.4 | 10.1 | 14736 | 1 | V52304 | Streptococcus pneu |
| 21 | 32.2 | 10.1 | 400 | 1 | V87584 | EST clone DY307. N |
| 22 | 32 | 10.0 | 4093 | 1 | Q49264 | ced-4. Long-distan |
| 23 | 32 | 10.0 | 1457 | 1 | V15409 | Methicillin resist |
| 24 | 32 | 10.0 | 80595 | 1 | V83939 | HC-contig derived |
| 25 | 32 | 10.0 | 80240 | 1 | V83940 | NC-contig derived |
| 26 | 32 | 10.0 | 18359 | 1 | X20255 | Borrelia burgdorfe |
| 27 | 31.8 | 9.9 | 1920 | 1 | Q88066 | Bacillus lichenifo |
| 28 | 31.8 | 9.9 | 1920 | 1 | Q95031 | Bacillus lichenifo |
| 29 | 31.8 | 9.9 | 1920 | 1 | V02471 | Bacillus lichenifo |
| 30 | 31.8 | 9.9 | 110000 | 1 | V21209_10 | Continuation (11 o |
| 31 | 31.6 | 9.9 | 134525 | 1 | Q04525 | Total base sequenc |
| 32 | 31.6 | 9.9 | 5852 | 1 | Q11710 | Dictyostelium plas |
| 33 | 31.6 | 9.9 | 5596 | 1 | V83941 | Bacterial artifical |
| 34 | 31.4 | 9.8 | 2207 | 1 | N80748 | Sequence of a poly |
| 35 | 31.4 | 9.8 | 1753 | 1 | V18455 | Human granulocytic |
| 36 | 31.4 | 9.8 | 110000 | 1 | X20248_01 | Continuation (2 of |
| 37 | 31.4 | 9.8 | 110000 | 1 | X20248_02 | Continuation (3 of |
| 38 | 31.4 | 9.8 | 10715 | 1 | X20248_09 | Continuation (10 o |
| 39 | 31.2 | 9.8 | 4472 | 1 | T05406 | Human cell cycle g |
| 40 | 31 | 9.7 | 4059 | 1 | T08973 | CI gene. Transgeni |
| 41 | 31 | 9.7 | 110000 | 1 | V21209_13 | Continuation (14 o |
| 42 | 31 | 9.7 | 110000 | 1 | V21209_14 | Continuation (15 o |
| 43 | 31 | 9.7 | 2025 | 1 | V54207 | Nucleotide sequenc |

ALIGNMENTS

| | | | | | | | |
|---|----|----|-----|------|---|--------|--------------------|
| c | 44 | 31 | 9.7 | 3072 | 1 | V54163 | Nucleotide sequenc |
| | 45 | 31 | 9.7 | 4059 | 1 | X15630 | Nucleotide sequenc |

RESULT 1
ID V58694 standard; CDNA; 320 BP.
AC V58694;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone P5105.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 134; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 320 BP; 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|--------|---------|---|-----|
| QY | 1 | CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACAACCAATAGGACAAAGAA | 60 |
| DB | 1 | CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACAACCAATAGGACAAAGAA | 60 |
| QY | 61 | AAAGTTGTGACATGTAGAGGAGTGTGACCCCTTACTCCCATCAAAAAAAT | 120 |
| DB | 61 | AAAGTTGTGACATGTAGAGGAGTGTGACCCCTTACTCCCATCAAAAAAAT | 120 |
| QY | 121 | GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAA | 180 |
| DB | 121 | GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAA | 180 |
| QY | 181 | GAGAAATACTACTTCTCTRAATGGAAGCCCTTAAAGGTCTTGATACTGAAGAC | 240 |
| DB | 181 | GAGAAATACTACTTCTCTRAATGGAAGCCCTTAAAGGTCTTGATACTGAAGAC | 240 |
| QY | 241 | AAATGCGCCGTCATCCCTTARAAGTGCATGATGGACACGTAACGTTCAGT | 300 |
| DB | 241 | AAATGCGCCGTCATCCCTTARAAGTGCATGATGGACACGTAACGTTCAGT | 300 |
| QY | 301 | TTTARACTCMGCATTGTGAC 320 | |
| DB | 301 | TTTARACTCMGCATTGTGAC 320 | |
| RESULT | 2 | | |
| ID | V61269 | standard; CDNA; 320 BP. | |
| AC | V61269; | | |

DT 06-JAN-1999 (first entry)
DE CDNA sequence of prostate tumour clone P510S.
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 124; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 320 BP; 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACCAACCAATAGAACACAAAGAA 60
Db 1 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACCAACCAATAGAACACAAAGAA 60
OY 61 AAAAGTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAAT 120
Db 61 AAAAGTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAAT 120
OY 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTCTAAAGAGAGAA 180
Db 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTCTAAAGAGAGAA 180
OY 181 GAGAAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTCTTGATCTGAAGGACAC 240
Db 181 GAGAAAATACTACTTCTCRAAATGGAAGCCCTTAAAGGTCTTGATCTGAAGGACAC 240
OY 241 AAATGTGGCCGTCACCTCTTARAGTTGCATGACTTGACACGCTAAGTGTGACAGT 300
Db 241 AAATGTGGCCGTCACCTCTTARAGTTGCATGACTTGACACGCTAAGTGTGACAGT 300
OY 301 TTTARACTCMGCATGTGAC 320
Db 301 TTTARACTCMGCATGTGAC 320

RESULT 3

V21209_01/c

Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular chr
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |

| | | | |
|----|-----------|---------|---------|
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

Query Match 14.1%; Score 45; DB 1; Length 110000;
Best Local Similarity 48.2%; Pred. No. 0.0093;
Matches 120; Conservative 2; Mismatches 127; Indels 0; Gaps 0;

OY 5 TGAAGGCTTCTGTAGAAATAGTACAGTTACCAACCAATAGAACACAAAGAGAAA 64
Db 93590 TGAAGACACCTTAAGATGAAGAAATATTTAAACCAATATCCAAACATATATAAAA 93531
OY 65 GTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGAT 124
Db 93530 ATTGTGAAGTTTAAAGACAGATGAGATATGATACCTTCCAAAAAATGCTT 93471
OY 125 ACATGTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGAGAGAGA 184
Db 93470 TAATTGGCAGGCAATTGAGAACTATTAATTTTAAATAAAAAATCTAAATAAAAAACGAAT 93411
OY 185 AATACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATCTGAAGGACACAAAT 244
Db 93410 ATATGTTCCAACTTTAGACAGAAACCAATATAATAGTTTAAAGAGATGATGAAT 93351
OY 245 GTGGCCGTC 253
Db 93350 ATTACCTTC 93342

RESULT 4

V21209_15/c

Continuation (16 of 17) of V21209 from base 1500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

| WP | Fragment Name | Begin | End |
|----|---------------|---------|---------|
| WP | V21209_00 | 1 | 110000 |
| WP | V21209_01 | 100001 | 210000 |
| WP | V21209_02 | 200001 | 310000 |
| WP | V21209_03 | 300001 | 410000 |
| WP | V21209_04 | 400001 | 510000 |
| WP | V21209_05 | 500001 | 610000 |
| WP | V21209_06 | 600001 | 710000 |
| WP | V21209_07 | 700001 | 810000 |
| WP | V21209_08 | 800001 | 910000 |
| WP | V21209_09 | 900001 | 1010000 |
| WP | V21209_10 | 1000001 | 1110000 |
| WP | V21209_11 | 1100001 | 1210000 |
| WP | V21209_12 | 1200001 | 1310000 |
| WP | V21209_13 | 1300001 | 1410000 |
| WP | V21209_14 | 1400001 | 1510000 |
| WP | V21209_15 | 1500001 | 1610000 |
| WP | V21209_16 | 1600001 | 1664976 |

Query Match 11.6%; Score 37; DB 1; Length 110000;
Best Local Similarity 49.2%; Pred. No. 1.1;
Matches 94; Conservative 1; Mismatches 96; Indels 0; Gaps 0;

| | |
|----|--|
| OY | 56 AAAGAAAAGTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAA 115 |
| Db | 61937 AAAGAAAAGTTTAAATACTTAAAGGTTTATATATGAGTAGTACTTACCCTTAGAAAA 61878 |
| OY | 116 AAATGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTCTTAAAGAGA 175 |
| Db | 61877 ATATGTATAGAAAAGCTTAAATATTTGGGCGAGTTGAAGTATATATGTTGTAATGATT 61818 |
| OY | 176 AGGAAGAGAAATACTCTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATCTGAAG 235 |
| Db | 61817 GCCCTATTAAATAAGACCGTTTCGAATGAAAAATCGTAAATGTTGCTTTATGACT 61758 |
| OY | 236 GACACAAATGT 246 |
| Db | 61757 AAAAACACAGT 61747 |

RESULT 5

T72882 standard; cDNA; 19124 BP.

AC T72882; (first entry)

DE Plasmodium var-7 gene.

KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy; Duffy binding like gene; Duffy antigen binding protein; erythrocyte; DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response; Plasmodium vivax.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT exon 7317..15139

FT /tag= a

FT /number= 1

FT Intron 15140..16205

FT /tag= b

FT /number= 1

FT exon 16206..17552

FT /tag= c

FT /number= 2

FT /note= "no stop codon given"

PN WO9640766-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; U09508.

PR 07-JUN-1995; US-487826.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X; Wellens TE; WPI: 97-052231/05.

DR P-PSDB; W22475.

PT New malaria vaccines - contains cysteine-rich DBL family protein binding domains homologous domains of the Duffy and stalic acid binding proteins

PS Claim 4; Page 56-61; 96pp; English.

CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to the Duffy binding like (DBL) family of genes which have homology to the Duffy antigen binding protein (DABP) and stalic acid binding protein (SABP) conserved regions (see T72889 and T72888 respectively). The var family of genes modulate cytoadherence and antigenic variation of Plasmodium infected erythrocytes. SABP and the Duffy antigen binding protein (DABP) are soluble proteins that appear in the culture supernatant after infected erythrocytes release merozoites. DABP and SABP mediate the binding of merozoites and schizonts to the erythrocyte surface. These proteins are necessary for erythrocyte invasion by the parasite. This sequence can be used in the compositions of the invention. The compositions are for the treatment and prevention of malaria, and comprise either a nucleotide sequence or encoded polypeptide of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of genes having homology with conserved regions of DABP and SABP. The compositions are used for the treatment and prevention of malaria. They are also used in the preparation of vaccines for inducing a protective immune response in a mammal to Plasmodium merozoites (especially Plasmodium falciparum or Plasmodium vivax).

Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 11.4%; Score 36.6; DB 1; Length 19124;

Best Local Similarity 51.5%; Pred. No. 0.96;

Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 36 ACAACCAATAGCAACAACAAAAAGTTGTGACATTGTAGGAGGTGTGTACC 95

DB 402 ACAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAGA 461

OY 96 CCTTACTCCCCATCAAAAAAATGATACATGTTAAAGATARAAGGCAATATTTT 155

DB 462 CATACGTTCACTTATTTATTAATGATTTATACGATTAAACATATTTGAGATTATAAT 521

OY 156 ATCATATGTTCTAAGAAGAAGAAGAAATACTACTTTCT 198

Db 522 AATATAATTACATAGAAAGAGTTAGAATACATTTT 564

RESULT 6

T68041 standard; DNA; 1215 BP.

AC T68041;

DT 21-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein ORF 079p11807orf42.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT cds 1..1215

FT /tag= a

FT /note= "no stop codon given"

PN WO9640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaerd BL; WPI: 97-052306/05.

DR P-PSDB; W20788.

PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) - useful for vaccines to treat or prevent H. pylori infection, and to detect Helicobacter

PS Claim 9; Page 862; 1481pp; English.

CC This sequence encodes a H. pylori cytoplasmic protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (AICC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.

Sequence 1215 BP; 379 A; 254 C; 214 G; 368 T;

Query Match 11.1%; Score 35.6; DB 1; Length 1215;

Best Local Similarity 56.0%; Pred. No. 0.93;

Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

OY 128 TGGTTAAGGATARAAGGCAATATTTATCATATGTTCTTAAGAAGAAGAA 187

DB 857 TGGGTAATGGGCCAAGCCTGATTTATTTAGATTTTAAAGAAATGAAGAAAT 916

OY 188 TACTACTTTCRAAATGGAAGCCCTTAAGGTGCTTTGATCTGAAGACACAAA 243

DB 917 TCATCATTTTTCATGCGGAACCGCTTTAAAGCCCTTAAAGCCGATGGCTTAA 972

RESULT 7

T67711 standard; DNA; 717 BP.

AC T67711;

DT 17-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein ORF 719606.aa.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor;

KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT cds 1..717


```
FT      /*tag= a
FT      /note= "no stop codon given"
PN      WO9640893-A1.
PD      19-DEC-1996.
PF      06-JUN-1996; U09122.
PR      07-JUN-1995; US-487032.
PR      01-APR-1996; US-630405.
PA      (ASTR ) ASTRA AB.
PI      Berglindh OT, Smith D, Mellgaerd BL;
DR      WPI; 97-052306/05.
P-PSDB; W20567.
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
PS      Claim 9; Page 315-318; 1481pp; English.
CC      This sequence encodes a H. pylori cytoplasmic protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
SQ      Sequence 717 BP; 224 A; 127 C; 146 G; 220 T;
```

```
Query Match      11.1%; Score 35.6; DB 1; Length 717;
Best Local Similarity 56.0%; Pred. No. 0.82;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;
```

```
QY      128 TGGTTAAAGGATARAAGGGCAATATTTATCATATGTTCTAAAGAGAAGAGAAAA 187
      ||| ||| ||| : | ||| ||| ||| ||| ||| ||| ||| |||
Db      434 TGGGTAAATGGCCCAAGCCGTGATTTATTGTTAGATTTTAAAGAAAAATGAGAAAAAT 493

QY      188 TACTACTTCTCRAAATGGAAGCCCTTAAAGGTGCTTGATGACTGAAGACACAAA 243
      | | | | | | | | | | | | | | | | | | | | |
Db      494 TCATCATTTTTCATGCGGACCGCTTAAAGCCTTAAAGCGCATGCGCTTAAA 549
```

```
RESULT      8
V74949      V74949 standard; DNA; 1107 BP.
AC      V74949;
DT      16-MAR-1999 (first entry)
DE      Staphylococcus aureus contig SEQ ID #638.
KW      Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW      cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW      skin infection; surgical wound infection; scalded skin syndrome;
KW      toxic shock syndrome; ds.
OS      Staphylococcus aureus.
FH      Key
FT      misc_feature      Location/Qualifiers
FT      541..600
FT      /*tag= a
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
PN      EP-786519-A2.
PD      30-JUL-1997.
PF      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI      Rosen CA;
DR      WPI; 97-374922/35.
PT      Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT      stored on computer readable medium and used in the production of
```

```
PT      anti-S.aureus vaccines
PS      Claim 1; Page 1565-1566; 3271pp; English.
CC      This sequence represents one of 5191 staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S.aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S.aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S.aureus DNA sequences contained on the
CC      computer readable medium.
SQ      Sequence 1107 BP; 384 A; 134 C; 194 G; 329 T;
```

```
Query Match      10.9%; Score 35; DB 1; Length 1107;
Best Local Similarity 57.7%; Pred. No. 1.3;
Matches 56; Conservative 2; Mismatches 39; Indels 0; Gaps 0;
```

```
QY      108 TCAAAAAAAATGATATCATGCTTAAAGATARAAGGCAATATTTATCATATGTTCT 167
      | | | | | | | | | | | | | | | | | | | | |
Db      221 TAAAAATAAAAATGGGGAAGTATATGCTTAAACATTCAGACGTTANCATNCTTCT 280

QY      168 AAAAGAGAAGAGAGAAATACTACTTCTCRAAAT 204
      ||| | | | | | | | | | | | | | | |
Db      281 AAAAAGAGCAAAAAGCATATACCGTTTCGTAGAAT 317
```

```
RESULT      9
V17183      V17183 standard; DNA; 1811 BP.
AC      V17183;
DT      18-JUN-1998 (first entry)
DE      Insecticidal toxin subunit BrhTX-1(a)1.1 genomic sequence.
KW      insecticidal toxin; Bracon hebetor; insect control; pathogen;
KW      recombinant baculovirus; ds.
OS      Bracon hebetor.
PN      WO9744355-A1.
PD      27-NOV-1997.
PF      01-MAY-1997; G01205.
PR      22-MAY-1996; GB-010748.
PR      22-MAY-1996; GB-010687.
PR      22-MAY-1996; GB-010695.
PR      22-MAY-1996; GB-010697.
PR      22-MAY-1996; GB-010738.
PR      22-MAY-1996; GB-010739.
PA      (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA      (ZENE ) ZENECA LTD.
PI      Christian PD, Claudianos A, Daly A, Duncan RE, Suter M,
PI      Windass JD;
DR      WPI; 98-018430/02.
PT      New nucleic acid encoding a combination of insecticidal sub:unit(s)
PT      of wasp toxin - and related transformed cells, insect pathogens and
PT      combinations of proteins, useful as insecticides
PS      Claim 1; Pages 57-58; 84pp; English.
CC      This is a contiguous genomic sequence which encodes part of a spliced RNA
CC      that is related to the cDNA comprised by pBrhTX-1(a)1.1. This 1811 base
CC      pair spliced RNA derived from the Bracon hebetor genomic clone encodes
CC      at least two of the insecticidal toxin subunits shown in sequences
CC      W52124-W52128. The spliced RNA can hybridise with extension products
CC      prepared from a 564 base pair (V17145) template with 6 specified primers.
CC      A nucleic acid encoding at least one of the specified subunits can be
CC      modified so that mRNA instability motifs and/or fortuitous splice sites
CC      are removed, or insect-pest preference codons are used, so that
CC      expression of this nucleic acid in insect cells yields practically the
```


CC same protein as unmodified nucleic acid in its endogenous organism. The
CC nucleic acid encoding an insecticidal toxin subunit can be complementary
CC to a sequence that hybridises under specified conditions to any of
CC sequences shown in V17145 to V17149. Cells transformed with these
CC nucleic acids, organisms regenerated from these cells and pathogens
CC containing these nucleic acids and insecticidal compositions comprising
CC a combinations of the toxin subunits are all used for control of insects.
CC The nucleic acids are used to produce recombinant baculoviruses for
CC insect control.
SQ Sequence 1811 BP; 649 A; 259 C; 292 G; 611 T;

Query Match 10.8%; Score 34.6; DB 1; Length 1811;
Best Local Similarity 55.4%; Pred. No. 1.9;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 26 TAGTACAGTTACCAACATAGAACACAAAGAAAGTTGTGACATTGTAGTAGG 85
DB 207 TAAACTCTCTCAATACTTAAAAAAGATCTGAAAGCTGTCTAAGA 266
OY 86 AGTGTGACCCCTTACTCCCATCAAAAAAATGATACATGTTAAAGATARAAG 145
DB 267 AATGTCCAGTTTGGACGCAAAACAAACGAAGAAACATGTATTAATTGTAGG 326
OY 146 G 146
DB 327 G 327

RESULT 10
V74675
ID V74675 standard; DNA; 10813 BP.
AC V74675;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #364.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 301..360
FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 2101..2160
FT /tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 3901..3960
FT /tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 5701..5760
FT /tag- d
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 7501..7560
FT /tag- e
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 9301..9360
FT /tag- f

FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PE 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1248-1254; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 10813 BP; 3513 A; 1895 C; 1594 G; 3451 T;

Query Match 10.7%; Score 34.2; DB 1; Length 10813;
Best Local Similarity 48.6%; Pred. No. 3.6;
Matches 90; Conservative 1; Mismatches 94; Indels 0; Gaps 0;

OY 11 CTTCTGTAGAAATAGTACAGTTACACCAATAGAACACAAAGAAAGTTGT 70
DB 107 CTACATTTTGAATAATCCATTTTTATACATTCACACTTTATATGACACTTCAT 166
OY 71 GACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGATACATG 130
DB 167 TACAGTTACTTTTATTTGTGATTCCTTACATGTTTCTTAAAAAATTTGTTATCATTA 226
OY 131 TTAAGGATARAAGGCAATATTTTATCATATGTTCTTAAAGAGAGAGAAATAC 190
DB 227 TTAACGTTGAATTAAGAAAAAATTAATCTGGAGATAAAAATGGAATATAAAAAGATAC 286
OY 191 TACTT 195
DB 287 TAATT 291

RESULT 11
X20248_04
Continuation (5 of 10) of X20248 from base 400001 (Borrelia burgdorferi polynucleotid
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

| Query Match | 10.7% | Score 34.2 | DB 1 | Length 110000 |
|-----------------------|--|----------------|----------|---------------|
| Best Local Similarity | 47.8% | Pred. No. 6.1 | | |
| Matches 96 | Conservative 1 | Mismatches 104 | Indels 0 | Gaps 0 |
| QY 12 | TTCTGTAGAAATAGTACAGTTACAACCAATAGGACACACAAAAGTGTG | 71 | | |
| Db 39340 | TGCTTGAAATACATATTAAATAAGAAACACAAACAGAACTTTGAAACTCTGG | 39399 | | |
| QY 72 | ACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAAAAATGATACATGGT | 131 | | |
| Db 39400 | CTATTGTAGGAATTAATTTTATTTTATTTTATTAATAAATAAGTCCAGCATTTAG | 39459 | | |
| QY 132 | TAAAGCATARAAGGCCAATTTTATTCATATGTTCTAAAAAGAGAAGACAGAAAATACT | 191 | | |
| Db 39460 | TAAATTACTTAGTGGATTTTATTAACATTAAATATAATATGATTAAACAAAATTAAA | 39519 | | |
| QY 192 | ACTTCTCRAAATGGAAGCCC | 212 | | |
| Db 39520 | AATATCTCAAAATTTAAACTC | 39540 | | |

```

RESULT      12
X20249/c
ID X20249 standard; DNA; 116277 BP.
AC X20249;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #2.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease; epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis; infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO, PI White OR;
PI WPI; 99-081217/07.
DR New isolated Borrelia burgdorferi nucleic acids - used to develop PT products for the detection, diagnosis, characterisation, prevention and therapy of infections, particularly Lyme disease
PT Claim 1; Page 672-737; 1128pp; English.
PS X20248 to X20402 represent polynucleotide sequences isolated from CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for the detection, diagnosis, characterisation, prevention and therapy of CC Bb infections, e.g. Lyme disease. They can also be used for the production of biosynthetic products, e.g. enzymes. Borrelia belongs to a family of motile, spiral-shaped bacteria called Spirochetes. Spirochetes are pathogenic in humans and Borrelia causes epidemic and endemic relapsing fever, and Lyme borreliosis, more commonly known as CC Lyme disease.
CC Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T; SQ

Query Match          10.7%; Score 34.2; DB 1; Length 116277;
Best Local Similarity 47.8%; Pred. No. 6.2;
Matches 96; Conservative 1; Mismatches 104; Indels 0; Gaps

QY 12 TTCTGTAGAAATAGTACAGTTACAACCAATAGGAACAACAAGAAAGTTGCG 71
    ||||| | | ||| | | | | | | | | | | | | | | | | | | |
Db 115794 TGCTTGAAAATACATATTAAAAATAGAAAACACACAAACAGAACTTTGGAAAACCTTCGG 115735

QY 72 ACAITGTAGTAGGAGGTGTGACCCTTACTCCCACATCAAAAAAAAAAATGATACATGGT 131
    ||||||| | | | | | | | | | | | | | | | | | | | | | |
Db 115734 CTATTGTAGGAATTAATTTTTTTTATTTTATTTTATAATAATAAGTCCAGCATTAG 115675

```

| | | | |
|----|--------|--|--------|
| OY | 132 | TAAAGCATARAAGGCCAATATTTTATCATATGTCTTAAGAAGGAAGAGAATAACT | 191 |
| | | | |
| D8 | 115674 | TAATTACTTAGTGGAATTTTTTATACATTAAATATAATATGATTAAACAATAATTA | 115615 |
| | | | |
| OY | 192 | ACTTTCRAAATGGAGCCC | 212 |
| | | : | |
| D8 | 115614 | AATATCTCAAAAATTTAACTC | 115594 |

RESULT 13
 ID X20259/c
 AC X20259 standard; DNA; 9842 BP.
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #12.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 OS Borrelia burgdorferi.
 PN W09858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; U12764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053344.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
 PI White OR;
 DR WPI: 99-081217/07.
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1: Page 914-919; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC Sequence 9842 BP; 3671 A; 1321 C; 1151 G; 3699 T;

| | Query Match | Best Local Similarity | Score | DB 1; | Length | 9842; |
|----|-------------|--|-------|------------|--------|--------|
| | Matches | 97; Conservative | 3; | Mismatches | 108; | Indels |
| | | | | | | Gaps |
| | | | | | | 0; |
| QY | 99 | TACTCCCCATCAAAAAAAAAAATGATACATGGTTAAAGGATARAAGGCCAATATTTTATC | 158 | | | |
| Dd | 6937 | TATTGGCAATAATAATAAAAAATTAACCTTCATAAAAAACATAATGATCATATTTTATT | 6878 | | | |
| QY | 159 | AFATGTCTTAAAAAGAAGGAAGAGAAAATACTACTTTCTCRAAATGGAAGCCCTTAAG | 218 | | | |
| Dd | 6877 | AGATTTTAAGATAATAAGATATATATAATTGCATTATCTGAACCTTAAAAACATTACT | 6818 | | | |
| QY | 219 | GTCCTTGATACTGAAGGACACAATGTGGCCGTCATCCTCTTARAGTTGCATGACT | 278 | | | |
| Dd | 6817 | AAAAAGCAAAGAAAAACCACCTATTAGGAAGTCCGATACCGGGCGCGGAGATGCAGAAAT | 6758 | | | |
| QY | 279 | TGGACACGGTAACTGTTGCAGTTTARA | 306 | | | |
| Dd | 6757 | TGTAGATGATCCTGATGGCAGATAGAA | 6730 | | | |

| RESULT | 14 |
|---|-------------------------------|
| X20248_03 | |
| Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide) | |
| WP Sequence split into 10 fragments | LOCUS X20248 Accession X20248 |
| WP | |
| Fragment Name | Begin End |

WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

Query Match 10.6%; Score 33.8; DB 1; Length 110000;
Best Local Similarity 53.4%; Pred. No. 7.8;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 42 AATAGGACACAAAGAAAGTTGTGACATTTAGTAGGAGTGTGACCCCTTAC 101
DB 14419 AATAATCTCAACAAATAAGAAAGAACTATAGCAATAAAATGCTTAAATATCTCCCAA 14478
OY 102 TCCCATCAAAAAAATGATACATGTTAAGGATARAAGGCAATATTTATCATA 161
DB 14479 ACTTCACACATGAATAAATAATAATGTAAGAAAGCTTATATCCCATTAAGCCACAT 14538
OY 162 TGTCTTAAAGAG 174
DB 14539 CTTCAAAAATAG 14551

RESULT 15

T72883
ID T72883 standard; cDNA; 7295 BP.
AC T72883;
DT 12-SEP-1997 (first entry)
DE Plasmodium var-1 gene.
KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
OS Plasmodium vivax.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 517..7065
FT /*tag- a
FT /product- var-1
PN W09640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09508.
PR 07-JUN-1995; US-487826.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
PI WPI: 97-052231/05.
DR P-PSDB: W22476.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and stalic acid
PT binding proteins
PS Claim 4; Page 67-69; 96pp; English.
CC This sequence represents the var-1 gene of Plasmodium. Var-1 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and stalic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The

CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 7295 BP; 2866 A; 1067 C; 1455 G; 1907 T;

Query Match 10.4%; Score 33.2; DB 1; Length 7295;
Best Local Similarity 51.4%; Pred. No. 5.9;
Matches 74; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

OY 15 TTGTAGAAATAGTACAGTTACACCAATAGACACAAAGAAAGTTGTGACA 74
DB 4885 TGGCTAGAAATATTTTGAAGATTAATAAGATTAGAAAAAATAAGCTATGTACG 4944
OY 75 TTGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAATGATACATGTTAA 134
DB 4945 AAAAAGGAAGATGATGCAAAATGTATAAAGGTTGTATAGAAAAATGGTACAGAAAA 5004
OY 135 AGGATARAAGGCCAATATTTATC 158
DB 5005 ACGAAAGATGGCAAAAATAAAC 5028

Search completed: September 25, 1999, 09:58:52
Job time: 8183 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:30:14 ; Search time 165.01 Seconds
(without alignments)
177.786 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320
Sequence: 1 CCCCTGAGCGCTTCTTGTTA.....TTTARACTCMGCATTGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-------------------|-------------------|
| C 1 | 38.2 | 11.9 | 7218 | 2 | US-08-232-463-14 | Sequence 14, Appl |
| 2 | 34.4 | 10.8 | 2367 | 3 | US-08-441-139-3 | Sequence 3, Appl |
| 3 | 31.8 | 9.9 | 1920 | 3 | US-08-720-899-1 | Sequence 1, Appl |
| 4 | 31.8 | 9.9 | 1920 | 3 | US-08-459-610-1 | Sequence 1, Appl |
| 5 | 31.8 | 9.9 | 1920 | 4 | US-08-343-804-1 | Sequence 1, Appl |
| 6 | 31.6 | 9.9 | 5852 | 1 | US-07-867-106-2 | Sequence 2, Appl |
| C 7 | 31.6 | 9.7 | 4059 | 2 | US-08-485-139-1 | Sequence 1, Appl |
| C 8 | 30.6 | 9.6 | 4435 | 2 | US-08-308-872B-5 | Sequence 5, Appl |
| C 9 | 30.6 | 9.6 | 321 | 4 | US-08-520-678A-23 | Sequence 23, Appl |
| 10 | 30.4 | 9.5 | 9636 | 3 | US-08-323-170B-1 | Sequence 1, Appl |
| 11 | 30.2 | 9.4 | 19557 | 5 | PCT-US92-06300-1 | Sequence 1, Appl |
| 12 | 29.8 | 9.3 | 12124 | 1 | US-08-181-271A-36 | Sequence 36, Appl |
| 13 | 29.8 | 9.3 | 12124 | 2 | US-08-449-315-36 | Sequence 36, Appl |
| 14 | 29.8 | 9.3 | 12124 | 2 | US-08-444-803-36 | Sequence 36, Appl |
| 15 | 29.8 | 9.3 | 12124 | 2 | US-08-449-043-36 | Sequence 36, Appl |
| 16 | 29.8 | 9.3 | 12124 | 2 | US-08-456-265A-36 | Sequence 36, Appl |
| 17 | 29.8 | 9.3 | 12124 | 3 | US-08-455-416-36 | Sequence 36, Appl |
| 18 | 29.8 | 9.3 | 12124 | 3 | US-08-455-244-36 | Sequence 36, Appl |
| 19 | 29.8 | 9.3 | 12124 | 3 | US-08-454-876-36 | Sequence 36, Appl |
| 20 | 29.8 | 9.3 | 12124 | 4 | US-08-457-364-36 | Sequence 36, Appl |
| 21 | 29.8 | 9.3 | 12124 | 4 | US-08-456-262-36 | Sequence 36, Appl |
| 22 | 29.8 | 9.3 | 12124 | 4 | US-08-456-240-36 | Sequence 36, Appl |
| 23 | 29.8 | 9.3 | 12124 | 4 | US-08-455-736-36 | Sequence 36, Appl |
| 24 | 29.6 | 9.2 | 15108 | 1 | US-08-157-005-1 | Sequence 1, Appl |
| 25 | 29.6 | 9.2 | 15101 | 5 | PCT-US95-09927-14 | Sequence 14, Appl |
| 26 | 29.4 | 9.2 | 1824 | 1 | US-08-347-792-5 | Sequence 5, Appl |
| 27 | 29.4 | 9.2 | 1824 | 2 | US-08-431-357-5 | Sequence 5, Appl |
| 28 | 29.4 | 9.2 | 2277 | 3 | US-08-676-967-2 | Sequence 2, Appl |
| 29 | 29.4 | 9.2 | 2291 | 3 | US-08-355-888A-6 | Sequence 6, Appl |
| 30 | 29.4 | 9.2 | 2277 | 3 | US-08-676-974-2 | Sequence 2, Appl |
| 31 | 29.4 | 9.2 | 2606 | 4 | US-08-568-459A-7 | Sequence 7, Appl |
| 32 | 29.4 | 9.2 | 2991 | 4 | US-08-588-190-1 | Sequence 1, Appl |
| 33 | 29.4 | 9.2 | 2991 | 4 | US-08-693-697-6 | Sequence 6, Appl |
| 34 | 29.4 | 9.2 | 2880 | 4 | US-08-693-697-32 | Sequence 32, Appl |
| 35 | 29.4 | 9.2 | 2877 | 4 | US-08-693-697-35 | Sequence 35, Appl |
| 36 | 29.4 | 9.2 | 246240 | 4 | US-08-724-394A-20 | Sequence 20, Appl |
| 37 | 29.4 | 9.2 | 246240 | 4 | US-08-724-394A-21 | Sequence 21, Appl |

| | | | | | | |
|------|------|-----|--------|---|-------------------|-------------------|
| 38 | 29.4 | 9.2 | 246240 | 4 | US-08-724-394A-22 | Sequence 22, Appl |
| 39 | 29.4 | 9.2 | 2880 | 4 | US-08-588-526-1 | Sequence 1, Appl |
| 40 | 29.4 | 9.2 | 1824 | 5 | PCT-US95-15353-5 | Sequence 5, Appl |
| 41 | 29 | 9.1 | 5181 | 3 | US-08-257-073-10 | Sequence 10, Appl |
| 42 | 29 | 9.1 | 11832 | 4 | US-08-416-603-1 | Sequence 1, Appl |
| C 43 | 29 | 9.1 | 356 | 4 | US-08-520-678A-22 | Sequence 22, Appl |
| C 44 | 29 | 9.1 | 257 | 4 | US-08-520-678A-24 | Sequence 24, Appl |
| C 45 | 29 | 9.1 | 253 | 4 | US-08-520-678A-25 | Sequence 25, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9pt-Fls
US-08-232-463-14

Query Match 11.9%; Score 38.2; DB 2; Length 7218;
Best Local Similarity 8.7%; Pred. No. 0.033;
Matches 24; Conservative 138; Mismatches 113; Indels 0; Gaps 0;

QY 20 AGAAATAGTACGTTACACCAATAGCAACAAGAAAGATTGTGACATGTGA 79
Db 1289 RR 1230
QY 80 GTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATACATGTTAAGGAT 139

[illegible]

```

RESULT      2
US-08-441-139-3
; Sequence 3, Application US/08441139
; Patent No. 5773245
;
; GENERAL INFORMATION:
;
; APPLICANT: Wiltrop, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
;
; TELEX: 230 901 SANS UR
;
; INFORMATION FOR SEQ ID NO: 3:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2367 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: 251..2176
;
US-08-441-139-3

```

| | Query Match | 10.8%; | Score 34.4; | DB 3; | Length 2367; |
|---------|--|---------------|-----------------|---------|--------------|
| | Best Local Similarity | 64.1%; | Pred. No. 0.27; | | |
| Matches | 50; Conservative | 1; Mismatches | 27; Indels | 0; Gaps | 0; |
| QY | 112 AAAAAAATGATACATCGTTAAAGCATARAAGGCCAATAATTTCATCATGTCTTCAA | A | 171 | | |
| | | | | | |
| Dd | 2275 AAAGAATAATGCTCGATTGGTAAGGAAGTAACAATGATATATTTTATTAATGTCTCAAA | A | 2334 | | |

QY 172 GAGAAGGAGAGAAATA 189
| | | | | | | |
Db 2335 AAAAAAAAAAAAAAAAAA 2352

```

RESULT      3
US-08-720-899-1
; Sequence 1, Application US/08720899
; Patent No. 5753460
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Svendsen, Allan
; APPLICANT: Thellersen, Marianne
; APPLICANT: Van der Zee, Pia
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57534600 No. 5753460disk of No. 5753460th America, Inc.
; STREET: 405 Lexington Avenue, 64th floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,899
; FILING DATE: 10-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/343,804
; FILING DATE: 22-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4054.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..1872
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 334..420
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 421..1869
US-08-720-899-1

```

| | Query Match | 9.9% | Score 31.8; | DB 3; | Length 1920; |
|----|--|-----------------|----------------|-----------|--------------|
| | Best Local Similarity | 56.8%; | Pred. No. 1.4; | | |
| | Matches 54; | Conservative 2; | Mismatches 39; | Indels 0; | Gaps |
| QY | 110 AAAAAAAAAATGCATACATGGTTAAAGCATARAAGGCCAATATTTCATATGTCTTCAA | 165 | | | |
| | | | | | |
| Dd | 187 AGAAGCAGAGAGCGCTATGCAATAAATGAGTAGAAGCCCATATCGCGCCTTTCTTTGG | 246 | | | |
| | | | | | |
| QY | 170 AAGAGAAGGAAGAGAAAAATACTACTTTCRCRAAT | 204 | | | |
| | : | | | | |
| Dd | 247 AAGAAAATATAGGGAATAATGCTACTTGTTAAAAAT | 281 | | | |


```

RESULT      4
US-08-459-610-1
; Sequence 1, Application US/08459610
; Patent No. 5801043
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; APPLICANT: Svendsen, Allan
; APPLICANT: Thellersen, Marianne
; APPLICANT: Van der Zee, Pia
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5801043o No. 5801043disk of No. 5801043th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,610
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/343,804
; FILING DATE: 22-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4054.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..1872
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 334..420
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 421..1869
;
US-08-459-610-1

Query Match          9.9%; Score 31.8; DB 3; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches    54; Conservative   2; Mismatches    39; Indels    0; Gaps    0;

OY  110 AAAAAAAAAATGCATCATGCTTAAAGGATARAAGCGCAATTATTATCATATGTTCATA 169
     ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  187 AGAAGCAGAGAGCGCTATGTGAATAATGAGTAGAGAAGCGCCATATCGCGCTTTCTTTGG 246

OY  170 AAAGAAGAGAAGAAAATACTACTTCTCRAAAT 204
     |||| | | | | | | | | | | | | | : ||||
DB  247 AAGAAAATATAGGGAATGCTACTTGTTAAAAAT 281

```

```

: Patent No. 5830837
: GENERAL INFORMATION:
: APPLICANT: Bisgaard-Frantzen, Henrik
: APPLICANT: Borchert, Torben Vedel
: APPLICANT: Svendsen, Allan
: APPLICANT: Thellersen, Marianne
: APPLICANT: van der Zee, Pia
: TITLE OF INVENTION: AMYLASE VARIANTS
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NO. 58308370 No. 5830837disk of NO. 5830837th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/343,804
: FILING DATE: 22-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lowney Dr., Karen A.
: REGISTRATION NUMBER: 31,274
: REFERENCE/DOCKET NUMBER: 4054.214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 334..1872
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 334..420
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 421..1869
: US-08-343-804-1

Query Match          9.9%; Score 31.8; DB 4; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches 54; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

Oy 110 AAAAAAAAAATGATACATGGTTAAAGCATARAAGGCCAATATTTATCATATGTTCTAA 169
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 AGAAGCAGAGAGGCTATTGAATTAATGAGTAGAAGCGGCATATCGCGCTTTCTTTTGG 246

Oy 170 AAGAGAAGAGAGAAATACTACTTCTTCRAAAT 204
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 AAGAAAATATAGGAAAATGTTACTTGTAAAAAT 281

RESULT 6
US-07-867-106-2
: Sequence 2, Application US/07867106
: Patent No. 5389526
: GENERAL INFORMATION:
: APPLICANT: Stæde, Martin B
: APPLICANT: Chang, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium

```


NAME/KEY: -
LOCATION: 1078..2134.
OTHER INFORMATION: /label=C1
OTHER INFORMATION: /note="coding region of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 2135..2430
OTHER INFORMATION: /label=3'C1
OTHER INFORMATION: /note="region containing polyadenylation signal"
OTHER INFORMATION: of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 1033..1038
OTHER INFORMATION: /label=TATA-Box
FEATURE:
NAME/KEY: -
LOCATION: 1061..1062
OTHER INFORMATION: /label=transcript-init
OTHER INFORMATION: /note="transcription initiation site"
FEATURE:
NAME/KEY: Intron
LOCATION: 1211..1299
FEATURE:
NAME/KEY: Intron
LOCATION: 1430..1575
FEATURE:
NAME/KEY: -
LOCATION: 935..939
OTHER INFORMATION: /label=C1-S
OTHER INFORMATION: /note="TGCAG sequence (in C1 gene) which in the
OTHER INFORMATION: C1-S sequence is changed to TTAGG"
US-08-485-139-1

Query Match 9.7%; Score 31; DB 4; Length 4059;
Best Local Similarity 61.7%; Pred. No. 3.1;
Matches 66; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

QY 18 TTAGAAATAGTACAGTGTACCAATAGGACACAAAAAGTTGTGACATTG 77
DB 3636 TTATTAATGACACAAATTCATCATATAGAGACAA-AATTAAGAGAAAGTTGTGAGATCA 3578
QY 78 TAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGGAT 124
DB 3577 AGTTCTAAATAGTCCCATGAAGTCAAACTTATTAATAAAGATAGAT 3531

RESULT 8

US-08-308-872B-5/C
Sequence 5, Application US/08308872B
Patent No. 5661006
GENERAL INFORMATION:
APPLICANT: BROWN, Thomas David Kay
APPLICANT: HORSBURGH, Brian Colin
TITLE OF INVENTION: CANINE CORONAVIRUS SUBUNIT VACCINE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Akzo No. 5661006el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,872B
FILING DATE: 19-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,846

FILING DATE: 05-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,641
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91.303.737.0
FILING DATE: 25-APR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4435 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Canine corona virus
STRAIN: CCV-V54
FEATURE:
NAME/KEY: CDS
LOCATION: 60..4418
OTHER INFORMATION: /label=CCV-C54_Spikegene
US-08-308-872B-5

Query Match 9.6%; Score 30.6; DB 2; Length 4435;
Best Local Similarity 47.5%; Pred. No. 4.1;
Matches 87; Conservative 1; Mismatches 95; Indels 0; Gaps 0;

QY 2 CCCTGAAGGCTCTTGTAGAAATAGTACAGTTACCAACCAATAGGACACAAAAAGAA 61
DB 429 CTCTATAGGCTGATATGTAACAATGATACACAGGACTACCATGATGTAAGTA 370
QY 62 AAAGTTGTGACATTTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATG 121
DB 369 AAGCTTACCTCTTGCATGCGAGTACTATTCGCATAGCTTCCATATCAAAATAAATG 310
QY 122 GATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTTAAAGAGAAGAG 181
DB 309 CATGATGTTACTAAATAATGTAAGCTGTGAGTGTGCTGTTCTGGAACAGTTGACC 250
QY 182 AGA 184
DB 249 ACA 247

RESULT 9

US-08-520-678A-23/C
Sequence 23, Application US/08520678A
Patent No. 5874565
GENERAL INFORMATION:
APPLICANT: Rice, Charles M.
APPLICANT: Kolykhalov, Alexander A.
TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Howell & Haferkamp, L.C.
STREET: 7733 Forsyth Blvd., Suite 1400
CITY: St. Louis
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30


```

1  CURRENT APPLICATION DATA:
2  APPLICATION NUMBER:  US/08/520,678A
3  FILING DATE:
4  CLASSIFICATION:  536
5  ATTORNEY/AGENT INFORMATION:
6  NAME:  Henderson, Melodie W.
7  REGISTRATION NUMBER:  37,848
8  REFERENCE/DOCKET NUMBER:  6029-6836
9  TELECOMMUNICATION INFORMATION:
10 TELEPHONE:  314-727-5188
11 TELEFAX:  314-727-6092
12
13 TELEX:
14
15 INFORMATION FOR SEQ ID NO:  23:
16
17 SEQUENCE CHARACTERISTICS:
18     LENGTH:  321 base pairs
19     TYPE:  nucleic acid
20     STRANDEDNESS:  single
21     TOPOLOGY:  linear
22
23     MOLECULE TYPE:  DNA (genomic)
24
25     US-08-520-678A-23

```

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 9.68; | Score 30.6; | DB 4; | Length 321; |
| Best Local Similarity | 60.08; | Pred. No. 1.6; | | |
| Matches 51; Conservative | 0; | Mismatches 34; | Indels 0; | Gaps 0; |

```

OY  20 AGAAATAGTACAGTTACACCAATAGGAACAACAAAAAGTTTGTCATTTGTA 79
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  170 AAAAAAAAAAAAAAAAAAGGGAACACAGAAATGGCCTAGA 111

```

```

QY      80 GTAGGAGTGTGTACCCCTTACTCC 104
          | | | | | | | | | | | |
Db     110 GGCCGGAGTGTTTACCCCAACCTTC 86

```

RESULT 10
US-08-323-

; Sequence 1, Application US/08323170E
; Patent No. 5733772

```

: GENERAL INFORMATION:
:
: APPLICANT: Williamson, Kim C.
:
: APPLICANT: Kaslow, David C.
:
: TITLE OF INVENTION: Cloning and Expression of Plasmodium
:
: TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pf5230
:
: NUMBER OF SEQUENCES: 4
:
: CORRESPONDENCE ADDRESS:
:
:

```

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/323,170E
 FILING DATE: 13-OCT-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/010,409
 FILING DATE: 29-JAN-1993
 ATTORNEY/AGENT INFORMATION:

```

;      REGISTRATION NUMBER: P-41,261
;      REFERENCE/DOCKET NUMBER: 015280-113100US
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (415) 576-0200
;      TELEFAX: (415) 576-0300
;      INFORMATION FOR SEQ ID NO: 1:
;      SEQUENCE CHARACTERISTICS:
;

```

```

; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
;
US-08-323-170B-1

```

| | | | | |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match | 9.58; | Score 30.4; | DB 3; | Length 9636; |
| Best Local Similarity | 52.58; | Pred. No. 6.2; | | |
| Matches 64; | Conservative 1; | Mismatches 57; | Indels 0; | Gaps 0; |

QY 108 TCAAAAAAAAAATGGATACATGCTTAAAGCATBAAGGCCAATATTTCATCATATGTTC 167
| | | | | : | | | | |
Db 4379 TCAAAAACAATGTGAAAAATCAATTAGAGAGAAAAGTTATATATCAATTTACTGTTCA 4438

Qy 168 AAAAGAGAGAGAGAGAAAAATACTACTTCTTCRAATGGAAGCCCTTAAAGGTGCTTTGA 227
 ||||| | | | | | | | | | | | | | | | |
 Db 4439 AAAAGAGACGAGAAAGTAAAGGTGTTTCATTGTGATTAATGAACATGCACATATGTTTACT 4498

| | | | |
|----|------|----|------|
| QY | 228 | TA | 229 |
| | | 11 | |
| Db | 4499 | TA | 4500 |

RESULT 11

PCT-US92-06300-1
; Sequence 1, Application PC/TUS92063000

```

; GENERAL INFORMATION:
; APPLICANT: Hurwitz, David R
; APPLICANT: Nathan, Margret
; APPLICANT: Shanl, Moshe
; TITLE OF INVENTION: Transgenic Protein Production
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
;

```

ADDRESS: Rhone-Poulenc Rorer, Inc
STREET: 500 Virginia Ave., Bldg. 3A
CITY: Ft. Washington
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19034

```

:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
:

```

APPLICATION NUMBER: PCT/US92/06300
 FILING DATE: 19920730
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Goodman, Rosanne
 REGISTRATION NUMBER: 52,534
 REFERENCE NUMBER: A0856-US
 TELECOMMUNICATION INFORMATION:

```

? TELEPHONE: (215) 962-4130
?
? TELEFAX: (215) 962-4107
?
? INFORMATION FOR SEQ ID NO: 1:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 19557 base pairs
?
? TYPE: NUCLEIC ACID
?
? STRANDEDNESS: double

```

```

; .      TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       PUBLICATION INFORMATION:
;       AUTHORS: Minghetti, P P
;       AUTHORS: Rufner, D E
;       AUTHORS: Ruang, W.-J.

```


APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericca C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/444,803
 FILING DATE: 19-MAY-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT 1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott

```

1
2      REGISTRATION NUMBER: 36,129
3      REFERENCE/DOCKET NUMBER: S-19835/P1/CGC 1727
4      TELECOMMUNICATION INFORMATION:
5          TELEPHONE: (919)541-8614
6          TELEFAX: (919)541-8689
7      INFORMATION FOR SEO ID NO: 36:
8          SEQUENCE CHARACTERISTICS:
9              LENGTH: 12124 base pairs
10             TYPE: nucleic acid
11             STRANDEDNESS: single
12             TOPOLOGY: linear
13             MOLECULE TYPE: DNA (genomic)
14             ORIGINAL SOURCE:
15             ORGANISM: Cucumis sativus
16             INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
17             IMMEDIATE SOURCE:
18             CLONE: pbscucchrct5
19
20 US-08-444-803-36

```

| | | | | |
|--------------------------|--------|----------------|-----------|---------------|
| Query Match | 9.38; | Score 29.8; | DB 2; | Length 12124; |
| Best Local Similarity | 49.78; | Pred. No. 10; | | |
| Matches 73; Conservative | 1; | Mismatches 73; | Indels 0; | Gaps 0; |

Oy 49 ACACACAAAAAGATTGTGACATTTAGTAGGGAGTGTACCCTTACTCCCAT 108
||| ||||| |||| | |||| | |||
Db 10697 ACAGCAAAAAAAAAATTGCCAAGSTTGAGTACCATTTCATAGATGCATATATAT 107

QY 109 CAAAAAATGCATACATGTTAAGGATARAAGGCCATATTTTATCATATGTTCTA 168

Db 10757 ATATATATATATATATTCATCAATATAAAAAATGAAATCTTTAAATAATATACA 10816

QY 169 AAAGAGAGGAGAGAAATACTACTT 195

Db 10817 GAGATGAAATTAACTAAATAGTAATT 10843

RESULT 15

; Sequence 36, Application US/08449043

Patent No. 5689044

GENERAL INFORMATION:

APPLICANT: Ryals, John A.

APPLICANT: Alexander, Danny C.

APPLICANT: Beck, James J.

APPLICANT: Duesing, John H.

APPLICANT: Friedrich, Leslie B

APPLICANT: Goodman, Robert M.

APPLICANT: HATHIS, CHRISTIAN

APPLICANT: Meins, Jr., Frederick
 ADDRESS: Montara 21400

APPLICANT: MOYER, ALICE
 APPLICANT: MOYER, MARY B

APPLICANT: Jean-Marc

APPLICANT: Bayne, George B

APPLICANT: Spierston
AFFIDICANT: Layne, George B.

APPLICANT: Stinson, Jeffrey B.

APPLICANT: Uknes, Scott J.

APPLICANT: Ward, Eric R.

APPLICANT: Williams, Shericca C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: U

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS 4.0C

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
IMMEDIATE SOURCE:
CLONE: pBscucchrht5
US-08-449-043-36

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;

Matches 73; Conservative 1; Mismatches 73; Indels 0; Gaps 0;
QY 49 ACACACAAAAAGTTGTGACATGTAGTAGGAGTGTTGTAACCCCTACTCCCAAT 108
||| ||||| ||||| ||| | ||||| ||| ||| |||
Db 10697 ACAGCAAAAAAATTCGAAGGGTGAAGTACCATGTGTCATAGATGCAATATATAT 10756
QY 109 CAAAAAATGATGATCATGTTAAAGCATRAAGGCAATATTTATCATATGTTCTA 168
| | | | | | | | | | | | | | | | | | | | |
Db 10757 ATATATATATATATATATATTCATCAATATAAAAAATTGAAAAATCTTAAATAATATATACA 10816
QY 169 AAGAGAGAGAGAGAGAAATACTACTT 195
| | | | | | | | | | | | | | | | | | | | |
Db 10817 GAGATGAATTAACATAATAGTAATT 10843

Search completed: September 25, 1999, 07:30:22
Job time: 3459 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:37 : Search time 181.29 Seconds
(without alignments)
348.487 Million cell updates/sec

Title: US-09-030-606-224
Perfect score: 320
Sequence: 1 CCCCTGAGGCTTCTGTGTA.....TTTARACTCMGCATGTGAC 320

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database : EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 202.2 | 63.2 | 512 | 21 R20590 | R20590 yg49d06.r1 |
| 2 | 201.2 | 62.9 | 285 | 40 Z45397 | Z45397 HSC20D041 n |
| 3 | 165.4 | 51.7 | 342 | 32 AA334754 | AA334754 EST39036 |
| 4 | 152 | 47.5 | 177 | 25 N56292 | N56292 JJ5699F Hum |
| 5 | 136.8 | 42.8 | 315 | 32 AA373221 | AA373221 EST85227 |
| 6 | 40.4 | 12.6 | 214 | 43 AU033864 | AU033864 AU033864 |
| 7 | 38.4 | 12.0 | 700 | 48 AI557583 | AI557583 pt2.1-1.D |
| 8 | 37.2 | 11.6 | 310 | 40 C93124 | C93124 C93124 Dict |
| 9 | 36.8 | 11.5 | 249 | 41 AI054289 | AI054289 q176c04.x |
| 10 | 36.8 | 11.5 | 269 | 42 AI144030 | AI144030 q163a08.x |
| 11 | 36.6 | 11.4 | 496 | 23 HI7024 | HI7024 ym39d10.s1 |
| 12 | 36.6 | 11.4 | 434 | 23 I47084 | I47084 SCMRAP171 C |
| 13 | 36.6 | 11.4 | 300 | 35 C57199 | C57199 C57199 Yuj1 |
| 14 | 36.6 | 11.4 | 300 | 36 C57822 | C57822 C57822 Yuj1 |
| 15 | 36 | 11.2 | 300 | 36 C59863 | C59863 C59863 Yuj1 |
| 16 | 35.8 | 11.2 | 354 | 44 AU038221 | AU038221 AU038221 |
| 17 | 35.2 | 11.0 | 504 | 35 AA550294 | AA550294 1423m3 gm |
| 18 | 35 | 10.9 | 587 | 50 AI667004 | AI667004 fc24d11.y |
| 19 | 34.8 | 10.9 | 471 | 46 AI404883 | AI404883 GH24733.5 |
| 20 | 34.6 | 10.8 | 432 | 36 AA604491 | AA604491 no73a09.s |
| 21 | 34.4 | 10.8 | 305 | 37 AA721414 | AA721414 n274b11.s |
| 22 | 34.4 | 10.8 | 520 | 39 AA825604 | AA825604 od57h06.s |
| 23 | 34.4 | 10.8 | 312 | 43 AI239707 | AI239707 gh43b02.x |
| 24 | 34.4 | 10.8 | 568 | 44 AU039059 | AU039059 AU039059 |
| 25 | 34.4 | 10.8 | 384 | 49 AI630952 | AI630952 ty99g12.x |
| 26 | 34.2 | 10.7 | 442 | 26 W47600 | W47600 zc35g07.r1 |
| 27 | 34.2 | 10.7 | 323 | 34 C24351 | C24351 C24351 Dict |
| 28 | 34.2 | 10.7 | 171 | 40 C90553 | C90553 C90553 Dict |
| 29 | 34.2 | 10.7 | 412 | 43 AI178240 | AI178240 EST221905 |
| 30 | 34.2 | 10.7 | 406 | 51 AI705892 | AI705892 UT-R-AC1- |
| 31 | 34 | 10.6 | 543 | 34 C24372 | C24372 C24372 Dict |
| 32 | 34 | 10.6 | 699 | 45 AU001859 | AU001859 AU001859 |
| 33 | 34 | 10.6 | 612 | 47 AI525038 | AI525038 promrna-5 |
| 34 | 34 | 10.6 | 373 | 48 AI582163 | AI582163 tq63b09.x |
| 35 | 34 | 10.6 | 477 | 53 HSM000181 | AI035888 Homo sapi |
| 36 | 33.8 | 10.6 | 464 | 20 Z32876 | Z32876 HHEA41N Altr |
| 37 | 33.8 | 10.6 | 466 | 22 R22059 | R22059 yh25g02.s1 |
| 38 | 33.8 | 10.6 | 583 | 25 N50519 | N50519 yy89b04.s1 |
| 39 | 33.8 | 10.6 | 388 | 25 N97950 | N97950 1711C3 czap |
| 40 | 33.8 | 10.6 | 383 | 25 N97969 | N97969 1731C3 czap |
| 41 | 33.8 | 10.6 | 427 | 31 AA279549 | AA279549 zs86b09.r |
| 42 | 33.8 | 10.6 | 565 | 46 AI407578 | AI407578 EST235868 |
| 43 | 33.8 | 10.6 | 353 | 47 AI522468 | AI522468 fd19b08.x |
| 44 | 33.8 | 10.6 | 312 | 49 AV020572 | AV020572 AV020572 |
| 45 | 33.6 | 10.5 | 602 | 34 AA532223 | AA532223 CpEST.501 |

ALIGNMENTS

RESULT 1
R20590 512 bp mRNA EST 18-APR-1995
LOCUS yg49d06.r1 Soares infant brain INIB Homo sapiens CDNA clone
DEFINITION IMAGE:36244 5', mRNA sequence.
ACCESSION R20590
NID 9775371
VERSION R20590.1 GI:775371

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 512)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Tresaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 321
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RPI
High quality sequence stop: 321.
Location/Qualifiers
1. 512
/organism="Homo sapiens"
/db_xref="GDB:408745"
/db_xref="taxon:9606"
/clone="IMAGE:36244"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lactimid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGAGAGATCGCGCGCCGACAGAAATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lactimid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 141 a 93 c 111 g 158 t 9 others
ORIGIN

Query Match 63.2%; Score 202.2; DB 21; Length 512;
Best Local Similarity 96.9%; Pred. No. 7.8e-38;
Matches 216; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

OY 1 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAGACAAAGAAAGA 60
DB 76 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAGACAAAGAAAGA 135
OY 61 AAAAGTTGTGACATGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAAT 120
DB 136 AAAAGTTGTGACATGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAAT 195
OY 121 GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAAGAGAGAA 180
DB 196 GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAAGAGAGAA 255
OY 181 GAGAAATACTACTTCTCRAAA--TGGAAGCCCTTAAAGGTG 221
DB 256 GAGAAATACTACTTCTCRAAAATGGGAAGCCCTTAAAGGG 298

RESULT 2
Z45397 245397 285 bp MRNA EST 01-MAY-1998
LOCUS

DEFINITION HSC20D041 normalized infant brain cDNA Homo sapiens cDNA clone
c-20d04, mRNA sequence.
ACCESSION 245397
NID 9574610
VERSION 245397.1 GI:574610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 285)
AUTHORS Anfiray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,
Sebastiani-Kabatchis, C. and Tessier, A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407250.

FEATURES
source
1. 285
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-20d04"
/clone_lib="normalized infant brain cDNA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: Lactimid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
Lactimid BA vector. Clone library from B.Soaes, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 101 a 44 c 52 g 86 t 2 others
ORIGIN

Query Match 62.9%; Score 201.2; DB 40; Length 285;
Best Local Similarity 98.0%; Pred. No. 1.4e-37;
Matches 200; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAGACAAAGAAAGA 60
DB 82 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACACCAATAGAGACAAAGAAAGA 141
OY 61 AAAAGTTGTGACATGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAAT 120
DB 142 AAAAGTTGTGACATGTAGTAGGAGTGTACCCCTTACTCCCATCAAAAAAAAT 201
OY 121 GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAAGAGAGAA 180
DB 202 GGATACATGTTAAAGGATARAAGGCAATATTTATCATATGTTCTAAAAGAGAGAA 261
OY 181 GAGAAATACTACTTCTCRAAAT 204
DB 262 GAGAAATACTACTTCTCRAAAT 285


```
RESULT      3
AA334754      342 bp      mRNA      EST      21-APR-1997
LOCUS      EST39036 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION      AA334754
ACCESSION      91987018
NID      AA334754.1 GI:1987018
VERSION      EST.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 342)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL      Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE      96026280
COMMENT      On May 9, 1995 this sequence version replaced gi:802598.
Other_ESTs: THC168716
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES
source
1..342
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):136461"
/db_xref="taxon:9606"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
/note="Vector: pBluescript SK-; site_1: EcoRI; site_2:
XhoI"
BASE COUNT      93 a      66 c      71 g      109 t      3 others
ORIGIN
Query Match      51.7%; Score 165.4; DB 32; Length 342;
Best Local Similarity 96.4%; Pred. No. 2.8e-29;
Matches 163; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
Db      121 CATGACTTGACACGGTACTGTGCAGTTTATAGACTCAGCATTTGTGAC 169
|||||
RESULT      4
N56292      177 bp      mRNA      EST      20-FEB-1996
LOCUS      J55699F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
DEFINITION      clone J55699 5', mRNA sequence.
ACCESSION      N56292
NID      91199140
VERSION      N56292.1 GI:1199140
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 177)
AUTHORS      Liew,C.C.
TITLE      cDNAs from fetal heart
JOURNAL      Unpublished (1995)
COMMENT      On May 18, 1995 this sequence version replaced gi:811148.
Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Seq primer: TCCCAAGATTGGCAGCAG.
FEATURES
source
1..177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="J55699"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
BASE COUNT      69 a      27 c      32 g      49 t
ORIGIN
Query Match      47.5%; Score 152; DB 25; Length 177;
Best Local Similarity 92.9%; Pred. No. 3.9e-26;
Matches 158; Conservative 1; Mismatches 11; Indels 0; Gaps 0;
```

```
QY      1 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACCAACCATAGGAACAACAAAAGA 60
|||||
Db      8 CCCCTGAGGCTTCTGTAGAAATAGTACAGTTACCAACCATAGGAACAACAAAAGA 67
61 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 120
|||||
Db      68 AAAAGTTGTGACATTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAT 127
121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAA 170
|||||
Db      128 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAA 177
|||||
RESULT      5
AA373221      315 bp      mRNA      EST      21-APR-1997
LOCUS      EST85227 HSCI172 cells 1 Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION      AA373221
ACCESSION      92025541
NID      AA373221.1 GI:2025541
VERSION      AA373221.1 GI:2025541
```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 315)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hung,J., Xu,C., Yu,G.L., Ruben,S.M.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M., and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On May 18, 1995 this sequence version replaced gi:811155.
Other_ESTs: THC168716

CONTACT: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tldb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers
1. 315
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):177582"
/db_xref="taxon:9606"
/clone_lib="HSCI72 cells I"
/cell_type="fibroblast"
/cell_line="HSCI72 (16PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 86 a 65 c 64 g 99 t 1 others
ORIGIN

Query Match 42.8%; Score 136.8; DB 32; Length 315;
Best Local Similarity 96.4%; Pred. No. 1.2e-22;
Matches 135; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 181 GAGAAATACTACTTCTCRAAATGAGCCCTTAAGTGCTTGTGATCTGAGGACAC 240
|||||
Db 1 GAGAAATACTACTTCTCAAAATGAGCCCTTAAGTGCTTGTGATCTGAGGACAC 60
QY 241 AAATGTGGCCGTCATCTCCTTARAGTTCATGACTGGACACGGTAACGTGTGACGT 300
|||||
Db 61 AAATGTGACCGTCATCTCCTTATAGAGTTCATGACTGGACACGGTAACGTGTGACGT 120
QY 301 TTARACTCMGCATGTGTGAC 320
|||||
Db 121 TTAGACTCAGCATGTGTGAC 140

RESULT 6

AU033864
LOCUS AU033864 214 bp mRNA EST 28-APR-1999
DEFINITION AU033864 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLB538, mRNA sequence.
ACCESSION AU033864
NID 93799288
VERSION AU033864.1 GI:3799288
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Dictyostelium discoideum.
AUTHORS Eukaryota; Dictyostelida; Dictyostelium.
1 (bases 1 to 214)
Mori,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,T., Ochiai,H. and Tanaka,Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
COMMENT On May 19, 1998 this sequence version replaced gi:2846208.

FEATURES
source Location/Qualifiers
1. 214
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLB538"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 149 a 22 c 12 g 31 t
ORIGIN

Query Match 12.6%; Score 40.4; DB 43; Length 214;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 92; Conservative 1; Mismatches 87; Indels 0; Gaps 0;

QY 10 GCTCTGTTAGAAATAGTACGTTCAACCAATAGCAACAACAAAAAGAAAGTTTG 69
|||
Db 25 GCCCTCTTAAATAAAAATTTGGTCACAAAAAAGAAAAAAGAAAAAGTAAG 84
QY 70 TGACATGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATACATG 129
|||
Db 85 AACCAACATTAATAAAAAAGAAAAAAGAAAAAAGAAAAAATTAATAAAAAAATT 144
QY 130 GTTAAAGATARAAGGCAATATTATCATATGTTCTAAAGAGAGAGAGAAATA 189
|||
Db 145 AATAAAGTAATATTAACAATAAATAATCAATCCCTTAATAAAAAAAGAAAAA 204

RESULT 7
A1557583 700 bp mRNA EST 23-MAR-1999
LOCUS A1557583
DEFINITION pt2.1-1.D10.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION A1557583
NID 94489946
VERSION A1557583.1 GI:4489946
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 700)
Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.


```
/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from
papillary serous ovarian carcinoma, CDNA made by oligo-dT
priming. Directionally cloned. Size-selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."
BASE COUNT      80 a      45 c      34 g      90 t
ORIGIN
QY  114 AAAAATGATACATGCTTAAGGATARAAGGCAATATTTATCATATGTTCTAAAGA 173
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  70 AAAATATCAAAACTGTTCAAGTCATATAAACACAAACATAAGTTTATTTAAAAAAA 129
QY  174 GAAGGAAGAGAAAATACTACTTCTTCRAAATGGAAGCCCTTAAGGTG 221
      ||| ||||| ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  130 AAAGAAAGAAAAAATAGTAATTCACAGAAATTCAGTGTTAGACGAG 177

RESULT 10
AI144030      269 bp      mRNA      EST      29-OCT-1998
LOCUS      q163a08.x1 NCI_CGAP_Ov26 Homo sapiens CDNA clone IMAGE:1861142 3',
DEFINITION      mRNA sequence.
ACCESSION      AI144030
NID      93665839
VERSION      AI144030.1 GI:3665839
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 269)
JOURNAL      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
      Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 344 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
Source
1.269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1861142"
/clone_lib="NCI_CGAP_Ov26"
/sex="female"
/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from
papillary serous ovarian carcinoma, CDNA made by oligo-dT
priming. Directionally cloned. Size-selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."
```

```
BASE COUNT      100 a      36 c      48 g      85 t
ORIGIN
Query Match      11.5%; Score 36.8; DB 42; Length 269;
Best Local Similarity 57.4%; Pred. No. 19;
Matches 62; Conservative 2; Mismatches 44; Indels 0; Gaps 0;
QY  114 AAAAATGATACATGCTTAAGGATARAAGGCAATATTTATCATATGTTCTAAAGA 173
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  182 AAAATATCAAAACTGTTCAAGTCATATAAACACAAACATAAGTTTATTTAAAAAAA 123
QY  174 GAAGGAAGAGAAAATACTACTTCTTCRAAATGGAAGCCCTTAAGGTG 221
      ||| ||||| ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  122 AAAGAAAGAAAAAATAGTAATTCACAGAAATTCAGTGTTAGACGAG 75

RESULT 11
HI7024      496 bp      mRNA      EST      29-JUN-1995
LOCUS      ym39d10.s1 Soares infant brain INIB Homo sapiens CDNA clone
DEFINITION      IMAGE:50782 3', mRNA sequence.
ACCESSION      HI7024
NID      9883264
VERSION      HI7024.1 GI:883264
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 496)
JOURNAL      Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
COMMENT      Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
      Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
      Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
      Wilson, R.
      The WashU-Merck EST Project
      Unpublished (1995)

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1544
High quality sequence stops: 368
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1544 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 368.
FEATURES
Source
1.496
/organism="Homo sapiens"
/db_xref="GDB:423592"
/db_xref="taxon:9606"
/clone="IMAGE:50782"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand CDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGAAGAATTCGGCGCCGACGAGATTTTCTTTTCTTTT 3'];
double-stranded CDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT      154 a      95 c      79 g      167 t      1 others
```


ORIGIN

| | | | | |
|--------------------------|--------|-----------------|-----------|-------------|
| Query Match | 11.48; | Score 36.6; | DB 23; | Length 496; |
| Best Local Similarity | 48.18; | Pred. No. 20; | | |
| Matches 99; Conservative | 1; | Mismatches 106; | Indels 0; | Gaps 0; |

OY 89 GTGTAACCCCTTACTCCCCCATCAAAAAAAAAAATGGATACATCGTTAAAGATTARAAGGCCA 148
 ||| |||| | | | | | | | | | | | | |
Dd 472 GTGAACCCCNCGCCTCCAATAAAAAATAAAAACCCTACTTTTGAGAATTCAAAGTA 413
 ||||| | | | | | | | | | | | | |
OY 149 ATATTTTCATATATGTCTTAAAGACAAGAGAAGAAAATACTACTTCTCRAAATGGAA 208
 | ||| | | | | | | | | | | | | : ||| | |
Dd 412 TACATGATCACAGTTTGCCCAAGAATTAGTAAAGACATTTTTATTTAAGTGAAGTGTAT 353
 | ||| | | | | | | | | | | | | | |
OY 209 GCCCTTAAGGTGCTTTGATACTGAAGACACAATGTGGCCGTCCATCCTCCTTARAG 268
 | | | | | | | | | | | | | | | | | |
Dd 352 CCTCAATTTATTGTGGAATTCCTCAGTCCGAATATCTGGAATAGGAACAGCTTTGTAG 293
 | | | | | | | | | | | | | | | | | |
OY 269 TTGCATGACTTGACACACGGTAAGTGT 294
 ||| | | | | | | | | | | | | | |
Dd 292 ATGCCTAACCATTAACCAACAATGAATGT 267

| RESULT | 12 |
|------------|--|
| L47084 | |
| LOCUS | L47084 |
| DEFINITION | 434 bp mRNA EST 18-MAR-1997 |
| | SCMRAP171 C2/25RY adult worm mini-library Schistosoma mansoni CDNA |
| | clone SMRAP171, mRNA sequence. |

| | | | | | |
|------------|-------|------|------|-------|----------|
| BASE COUNT | 142 a | 70 c | 92 g | 122 t | 8 others |
| ORIGIN | | | | | |

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 11.48; | Score 36.6; | DB 23; | Length 434; |
| Best Local Similarity | 59.48; | Pred. No. 20; | | |
| Matches 60; Conservative | 0; | Mismatches 41; | Indels 0; | Gaps 0; |

OY 54 AAAAGAAAAAGTTTGTGCACATTTGTAAGGGAGTGTAACCCTTACTCCCATCAAAA 113
+
Db 172 ATAAATAAAGACAATTTAACATTTTGGTAGAGCGGTATGTCGGCTATACTGAATGTATGAA 231

OY 114 AAAAATGCATCATGGTTAAAGCATABAAGGGCAATATTT 154

Db 232 AAAAGTNGTTACACGNCCTACACGATATATCTTACTTTT 272

| RESULT | 13 |
|------------|---|
| LOCUS | C57199 |
| DEFINITION | C57199 300 bp mRNA EST 22-SEP-1997 C57199 Yuj1 Kohara unpublished cDNA <i>Caenorhabditis elegans</i> CDNA clone yk300c9 3', mRNA sequence. |
| ACCESSION | C57199 |
| NID | 92415230 |
| VERSION | C57199.1 GI:2415230 |
| KEYWORDS | EST. |
| SOURCE | <i>Caenorhabditis elegans</i> . |
| ORGANISM | <i>Caenorhabditis elegans</i> Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; <i>Caenorhabditis</i> . 1 (bases 1 to 300) |
| REFERENCE | Kohara, Y., Motchashvili, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A. Expression map of the <i>C.elegans</i> genome Unpublished (1996) |
| TITLE | |
| JOURNAL | |
| COMMENT | On Sep 12, 1996 this sequence version replaced gi:1394822. |

| BASE COUNT ORIGIN | 120 a | 30 c | 67 g | 82 t | 1 others |
|----------------------|-------|------|------|------|----------|
|----------------------|-------|------|------|------|----------|

| | | | | |
|--------------------------|--------|----------------|-----------|-------------|
| Query Match | 11.48; | Score 36.6; | DB 35; | Length 300; |
| Best Local Similarity | 55.98; | Pred. No. 21; | | |
| Matches 66; Conservative | 1; | Mismatches 51; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|---|-----|
| 2y | 52 | ACAAAAAGAAAAGTTTGTGACATTTAGTAGGGAGTGTAACCCCTTACTCCCACATCA | 111 |
| | | | |
| Db | 145 | AAAAAGAGAAGATATTGTNATGAGTGGGTGGGGAGTGAACCGCTGATTTTGAGCAA | 204 |
| | | | |
| Qy | 112 | AAAAAAAAATGCATCATCGTTAAAGCATARAAGGCCAATATTTATCATATGTTCTAA | 169 |
| | | : | |
| Db | 205 | AATGAATATACATTAAGTGAAGGAAATATGATATATTAAGTAAAGTTTAA | 262 |

| RESULT | 14 |
|------------|---|
| C57822 | |
| LOCUS | C57822 300 bp mRNA |
| DEFINITION | C57822 Yuj1 Kohara unpublished cDNA <i>Caenorhabditis elegans</i> CDNA clone yk310a1 3', mRNA sequence. |

Wed-Sep-29 14:27:52 1999

us-09-030-606-224_1.rst

This Page Blank (usps)